Gibbs Fields with Multiple Pairwise Pixel Interactions for Texture Simulation and Segmentation

Georgy L. Gimel'Farb

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Gibbs Fields with Multiple Pairwise Pixel Interactions for Texture Simulation and Segmentation

Georgy L. Gimel'farb

N° 3202

July 1997

THÈME 3

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Gibbs Fields with Multiple Pairwise Pixel Interactions for Texture Simulation and Segmentation

Georgy L. Gimel'farb*

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet PASTIS

Rapport de recherche n°3202 — July 1997 — 68 pages

Abstract: Modelling of spatially homogeneous and piecewise-homogeneous image textures by novel Markov and non-Markov Gibbs random fields with multiple pairwise pixel interactions is briefly overviewed. These models allow for learning both the structure and strengths (Gibbs potentials) of the interactions from a given training sample. The learning is based on first analytic and then stochastic approximation of the maximum likelihood estimates (MLE) of the potentials. A novel learning approach, giving explicit, to scaling factors, estimates of the potentials, is outlined. It exploits the conditional MLE provided that the training sample may rank a feasible top place within the parent population in its total Gibbs energy. The models embed both simulation and segmentation of the grayscale piecewise-homogeneous textures into the same Bayesian framework exploiting a controllable simulated annealing to generate the desired texture or its region map. Experimental results in simulating and segmenting various natural textures are presented and discussed.

Key-words: Texture simulation, segmentation, Gibbs field, parameter estimation

(Résumé : tsvp)

* *on leave from* International Research and Training Center for Information Technologies and Systems of the National Academy of Sciences and Ministry of Education of Ukraine, prospect Glushkova 40, 252022 Kiev-22, Ukraine; *phone*: ++(380 44) 266 2569; *fax*: ++(380 44) 266 1570; *e-mail*: gimel@image.kiev.ua

Unité de recherche INRIA Sophia Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex (France)
Téléphone : 04 93 65 77 77 – Télécopie : 04 93 65 77 65

Champs de Gibbs avec interactions de paires multiples en simulation et segmentation de textures

Résumé : On présente une nouvelle modélisation d'images de textures homogènes globalement ou par morceaux (régions). Nous faisons pour cela appel à de nouveaux modèles de champs de Gibbs, markoviens ou non-markoviens, mettant en jeu des interactions de paires multiples. Ces modèles permettent l'apprentissage simultané de la structure des interactions et de la force des potentiels de Gibbs associés à partir d'un échantillon d'apprentissage initial. La méthode d'apprentissage est basée sur une approximation analytique, puis stochastique, de l'estimée au maximum de vraisemblance de ces potentiels de Gibbs. Une nouvelle méthode donnant l'estimation de ces potentiels (à un facteur d'échelle prés) est ainsi precisée. Elle exploite l'estimation du maximum de la vraisemblance conditionnelle, à condition que l'échantillon d'apprentissage ait des caractéristiques suffisamment proches de la population parente de la texture à estimer. Ces modèles replacent la simulation et la segmentation de textures homogenes par morceaux dans un même cadre bayesien et mettent en jeu un "recuit simulé controlable" afin de générer soit la texture désirée, soit la carte des labels associée. Des résultats expérimentaux en simulation et en segmentation de diverses textures naturelles sont présentés et discutés.

Mots-clé : simulation de texture, segmentation, champ de Gibbs, estimation de paramètre

1 Introduction: Structure and Texture

A great many works on image modelling in present computer vision and image processing are dealing with spatial image structures and image textures. Nonetheless, both these latter terms, as state, for instance, HARALICK and SHAPIRO [1992], are lacking formal and precise definitions. Informal human views can be found in most standard lexicons, say, OXFORD [1971], WEBSTER [1959, 1986], BARNHART and BARNHART [1990], etc.

The **structure** (from a Latin *structura* which means "to build, arrange") is defined in a broad sense as a mutual relation of the elements of the whole object or a fabric or framework of putting them together.

The **texture** (from a Latin *textura* which means "weaving", "web", or "structure") relates to a specific structure of visual or tactile surface characteristics of particular objects such as natural woven ones (fabrics, tissues, weaves, webs, paintings). In a broad sense, the texture defines also the structure or composition of an object with regard to its components.

Probably, most extensively both terms are elaborated in modern petrography: the *texture* describes smaller features of a rock depending upon a size, shape, arrangement, and distribution of the components but the *structure* relates to the larger features such as foldings, faults, crackings, etc. In other areas the texture is also referred mostly to small-scale surface features or fine structures of the objects and only in rare cases replaces directly the structure.

Despite both notions - the structure and the texture - are rather close in the meaning, the former is much more universal. One may consider the internal structure of a given texture, that is, study its particular elements and their arrangement. But, it seems rather odd to consider "a texture of a structure" (though this involves no contradictions).

1.1 Human view on image textures

Thus, for a human, the texture relates mostly to specific spatially homogeneous, to some extent, (micro)structures which are obtained by "weaving" of the object elements so that their arrangement matches, in a broad sense, visual (or tactile) features of a woven fabric. The term "image textures" refers both to grayscale, color, or multiband images of the natural textured surfaces and to simulated patterns which approach, within certain limits, these natural images. Imaging of the natural textures involve additional projective or orientation-scale geometric and specific linear or non-linear photometric transformations of the percei-

ved structures within a given texture type or example.

The texture is a somewhat more restricted notion than the structure itself because it is represented, mostly, by particular natural examples, say, granular, porphyritic, fluidal or other more specific types in petrography, or fine - coarse, close - loose, plain - twilled or ribbed textile fabrics, etc. Nevertheless, these human classifications are too fuzzy to form a basis for a formal definition of image textures in image processing and computer vision.

1.2 Computational view on image textures

As a result, present computational views on image textures either exploit particular image sets, selected as generally recognized *de facto* prototypes (like, for instance, the well-known photoalbum of BRODATZ [1966] and the recent digital collection of PICKARD et al. [1995]), or still involve informal qualitative features such as fineness, coarseness, smoothness, granulation, lineation, directionality, roughness, regularity, randomness, *etc*. Unfortunately, there are few attempts to associate these features, more or less easily perceived by a human, with computational image models used to describe and analyze the image textures.

The texture model involves basic gray level (or color, or multi-band) texture primitives that form texture elements, called *textons* by JULESZ [1981] or *texels* by HARALICK and SHAPIRO [1992], built from one or several primitives. Spatial interactions between the texels govern their spatial arrangement and particular signal values in each of the primitive. In this context, the interaction has no direct physical meaning and deals only with relative frequencies of particular spatial signal configurations formed by the texels: the more frequent the configuration, the stronger the interaction. Such interpretation of the interaction suggests that the basic property of the image to be a homogeneous texture is a *spatial self-similarity*, that is, a (statistical) repetitiveness of specific signal configurations, say, the texels with particular signal values, over the image. Of course, this definition is still informal but, at least, proposes a way of building the formal one which has to specify quantitatively what and how has to be repeated in the texture.

For a human, the spatial self-similarity, or repetitiveness, is quite definitive. But, it is very hard (if possible) to give a general constructive formal definition of it. An early self-similarity concept of CHETVERIKOV [1987] referred to a subimage (patch) of minimum size that can be considered as the texture and to a binary similarity relation between the patches. But, the ways how to find them in practice were lacking. More elaborated approach of ZALESNY [1994] involves distances between conditional probability distributions

of a particular function of the gray level configurations in the patches as a quantitative measure of the similarity. The patches are certain translation-invariant connected pixel subsets which are equivalent under a given group of transformations. But, in the general case, it is unclear how to search for these patches and functions for testing the spatial self-similarity of a given image.

To avoid the impracticable generalizations, we restrict herein the self-similarity concept to conditional probability distributions of image signals in the texels and to the pixels and pixel pairs as the (simplest) texels. Such a restriction allows for a precise definition of the spatial self-similarity in terms of quantitative parameters of particular Gibbs Random Fields (GRF) with multiple pairwise pixel interactions. These novel Gibbs models of spatially homogeneous and piecewise-homogeneous image textures as well as approaches how to learn their parameters were proposed recently by GIMEL'FARB [1996a] - [1996c]. The images are considered as samples of the GRF defined by a particular Gibbs Probability Distribution (GPD). The model parameters, that is, the parameters of the GPD, defining the GRF, specify a geometric structure and quantitative strengths of the pairwise interactions. The structure is represented by several families of the translation invariant pixel pairs. The interaction strength for each family is given by a particular Gibbs potential function of the image signals in the pixel pair, or texel. Basically, such a simplification leaves aside most of the textures because usually they have larger and more diverse texels and greater geometric differences between the self-similar parts. Nevertheless, the experiments show that a sufficient number of natural and artificial image textures are adequately modelled with these simple models.

1.3 Contents of the paper

Sections 2 and 3 give a brief overview of the Markov and non-Markov GRFs with multiple pairwise pixel interactions. Generalized image models with varying interaction structure are introduced to represent in more details the piecewise-homogeneous textures and their region maps. Learning of the model parameters is outlined in Section 4. Both the interaction structure and the potentials are learnt, from a given training sample, by using first analytic and then stochastic approximation of the unconditional Maximum Likelihood Estimate (MLE) of the potentials. These approximations exploit the histograms of the image signals and of the signal co-occurrences or differences which form a sufficient statistic for the models. Section 5 discusses a novel learning approach based on the conditional MLEs such that the training data sample may reach its highest rank within the parent population in its total Gibbs energy (GIMEL'FARB, SCHMIDT, and BRAUNMANDL [1997]). In this case the centered signal histograms represent the explicit, to scaling factors, MLE of the Gibbs

potentials. This allows to build more detailed Gibbs image models provided that the given training sample is sufficiently large to obtain consistent estimates of certain conditional probability distributions from the collected histograms. The required sample sizes are discussed in Section 6 where a model-based measure to check the texture self-similarity is also introduced. Section 7 presents results of simulating (generating) and segmenting various natural image textures and gives some conclusions.

1.4 Basic notation

Below, we use the following basic notation.

$\mathbf{R} = \{(m,n) : m = 0,\ldots,M-1;\ n = 0,\ldots,N-1\}$ is a finite arithmetic 2D lattice with $M \cdot N$ pixels supporting digital images. In most cases, the pixel will be denoted by a shorthand symbol $i \equiv (m,n)$.

$\mathbf{g} = \{g(i) : i \in \mathbf{R}\}$ denotes a digital grayscale image with gray levels $q = g(i) \in \mathbf{Q}$. The gray levels, measured in a metric scale, have a finite set $\mathbf{Q} = \{0,\ldots,q_{\max}\}$ of values. Usually, $q_{\max} = 15$ or 255. A gray level difference for a pixel pair (i,j) is denoted by $d = g(i) - g(j) \in \mathbf{D}$ where $\mathbf{D} = \{-q_{\max},\ldots,0,\ldots,q_{\max}\}$ is a finite set of the difference values.

$[q_{\min}(\mathbf{g}), q_{\max}(\mathbf{g})]$ is the gray range for the image $\mathbf{g}$. Here, $q_{\min}(\mathbf{g}) = \min\limits_{i\in\mathbf{R}}\{g(i)\}$ and $q_{\max}(\mathbf{g}) = \max\limits_{i\in\mathbf{R}}\{g(i)\}$ are the minimum and maximum gray values in the image $\mathbf{g}$.

$\mathbf{l} = \{l(i) : i \in \mathbf{R}\}$ denotes a digital region map, that is, the map of homogeneous regions in the lattice. Region labels $k = l(i) \in \mathbf{K}$, measured in a nominal scale, have a finite set $\mathbf{K} = \{0,\ldots k_{\max}\}$ of values.

$\mathbf{G}$ and $\mathbf{L}$ are the parent populations of the grayscale images and of the region maps, respectively.

$\mathbf{C}_a = \{(i,j) : i,j \in \mathbf{R};\ i - j = (\mu_a,\nu_a)\}$ denotes a family of translation invariant pixel pairs with a fixed shift between the pixels in the lattice.

$\mathbf{C} = \{\mathbf{C}_a : a \in \mathbf{A}\}$ is a given set of the families with indices a matching one-to-one with the corresponding shifts (μ_a,ν_a). Here, $\mathbf{A}$ is a set of the indices. The families $\mathbf{C}$ show a geometric structure of pairwise pixel interactions in the lattice in terms of fixed orientations of the pixel pairs $\varphi_a = \arctan(\mu_a/\nu_a)$ and distances $(\mu_a^2+\nu_a^2)^{\frac{1}{2}}$ between these pixels.

$\mathbf{H}_{\text{pix}}(\mathbf{g}) = \{H(q|\mathbf{g}) : q \in \mathbf{Q}\}$ denotes the Gray Level Histogram (GLH) collected over the lattice $\mathbf{R}$ for the grayscale image $\mathbf{g}$:

$$H(q|\mathbf{g}) = \sum_{i\in\mathbf{R}} \delta(q - g(i))$$

where $\delta(\ldots)$ is the Kronecker function.

$\mathbf{H}_{\mathrm{pix}}(\mathbf{g},\mathrm{l}) = \{H(q,k|\mathbf{g},\mathrm{l}) : q \in \mathbf{Q};\ k \in \mathbf{K}\}$ denotes the joint Gray Level and Region Label Histogram (GL/RLH) collected over the lattice $\mathbf{R}$ for a superimposed pair of the grayscale image $\mathbf{g}$ and the region map l:

$$H(q,k|\mathbf{g},\mathrm{l}) = \sum_{i\in\mathbf{R}} \delta(q-g(i))\cdot\delta(k-l(i)).$$

$\mathbf{H}_a(\mathbf{g}) = \{H_a(d|\mathbf{g}) : d \in \mathbf{D}\}$ denotes the Gray Level Difference Histogram (GLDH) collected over the family $\mathbf{C}_a$ for the grayscale image $\mathbf{g}$:

$$H_a(d|\mathbf{g}) = \sum_{(i,j)\in\mathbf{C}_a} \delta(d-(g(i)-g(j))).$$

$\mathbf{H}_a(\mathbf{g},\mathrm{l}) = \{H_{a,\alpha}(d,k|\mathbf{g},\mathrm{l}) : d \in \mathbf{D};\ k \in \mathbf{K};\ \alpha \in \{0,1\}\}$, denotes the joint Gray Level Difference and Region Label Co-incidence Histogram (GLD/RLCH) collected over the family $\mathbf{C}_a$ for the superimposed pair of the grayscale image $\mathbf{g}$ and the region map l:

$$H_{a,\alpha}(d,k|\mathbf{g},\mathrm{l}) = \sum_{(i,j)\in\mathbf{C}_a} \delta(d-(g(i)-g(j)))\cdot\delta(k-l(i))\cdot\delta(\alpha-\delta(l(i)-l(j))).$$

Here, α indicates the intra-region ($\alpha = 1$, or, $l(i) = l(j)$) and inter-region ($\alpha = 0$, or, $l(i) \neq l(j)$) pixel interactions.

The dot product of two vectors is denoted by $\bullet$.

2 Modelling Homogeneous Stochastic Textures

Probabilistic image models attach to each image a numerical tag, namely, a probability to receive this image in certain contexts. If the model is generative, it allows also to generate (simulate, or compute) the images having a given probability distribution. In other words, frequencies of the generated images approach the given probabilities in the limit when the number of generated images tends to infinity. In image modelling, one needs the following features of the probability distributions:

- (*i*) "Focussing" on a very small desired subset of the images from the parent population:

  image processing and analysis applications usually involve no more than hundreds, thousands, or, in rare cases, a few millions of digital grayscale images whereas the parent populations contain an incomparably greater number $|\mathbf{Q}|^{|\mathbf{R}|}$ of the samples, for instance, $10^{100}\ldots10^{10000}$ images even for rather small lattices $10\times10\ldots100\times100$ with only 10 gray levels per pixel.

- (*ii*) Learning capabilities, that is, possibilities to estimate quantitative model parameters from a given training sample or a subset of the samples.
- (*iii*) Moderate computational complexity of the image generation and of the parameter estimation:

  because most applications involve lattices of $100 \times 100 \ldots 1000 \times 1000$ pixels and even more, then the processing is computationally feasible only if the complexity is $O(|\mathbf{R}|)$ or less.
- (*iv*) Mutual compatibility between a joint probability distribution of the image signals and conditional probability distributions of the signal subsets in the image.

The GRFs as image models offer all these features. The models describe the images in terms of an explicit geometric structure and quantitative strengths of the pixel interactions. Herein, we restrict the consideration only to multiple translation invariant pairwise pixel interactions and, therefore, to a specific class of spatially homogeneous and piecewise homogeneous image textures, called *stochastic textures* in (GIMEL'FARB [1996a]). The stochastic texture has pixels and pixel pairs as the (primitive) texels so that it is specified by the structure and strengths of the translation invariant pairwise pixel interactions.

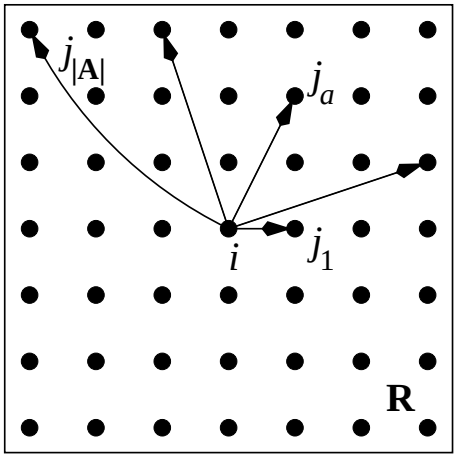


Figure 1: Pairwise interactions.

The interaction structure is represented usually by a *neighborhood graph* (see, for instance, BESAG [1974]). The pixels are the vertices of the graph and its arcs join the interac-

ting pixel pairs called the *neighbors*. In the Gibbs models under consideration, a particular interaction is given by a translation invariant family $\mathbf{C}_a$ of the pixel pairs. Each pair has the same spatial shift (μ_a, ν_a) between the pixels and differs from other pairs only by its absolute position in the lattice. The inter-pixel shifts specifying the families can be arbitrary and depends on a texture type to be modelled (see, for instance, Figure 1). The GPD is factored over these families, each factor being an exponential function of a (Gibbs) *potential*. Each family has its own potential that depends on the signals in the pixel pair. The potential value defines quantitatively the interaction strength in this pair. In the Markov/Gibbs models these pairs are *cliques*, or complete subgraphs of the neighborhood graph (BESAG [1974], HASSNER and SKLANSKY [1980], GEMAN and GEMAN [1984]). In the simple non-Markov Gibbs models introduced in (GIMEL'FARB [1996b]) these pairs are not the cliques because the neigborhood of each pixel coincides, in principle, with the total lattice. But, they represent still most significant part of the "lattice-wide" interactions. For brevity, we refer below to them by the same terms "cliques" and "clique families".

2.1 Traditional Markov/Gibbs models

The Gibbs models were used first in statistical physics to describe the behavior of large systems of interacting particles (ISIHARA [1971]). Here, the interaction structures and strengths have obvious and constructive physical parallels. But, in the context of image textures it is necessary to define the exact meaning of these notions because the straightforward analogies with physical interactions between the particles, with energies of the particle systems, with temperatures of the systems, etc., are mostly misleading.

The *pixel interaction* means almost no more that the gray levels in each particular clique do not take purely random values. In other words, their probabilities (or relative frequencies) differ from those of the Independent Random Field (IRF): the more probable (that is, more frequent) a particular spatial signal configuration in the texture, the higher the interaction between these signals, or, in terms of the Gibbs potentials, the higher the positive potential value for this configuration. Therefore, the only image feature to be linked to the pixel interactions is the frequency or probability of the signal configurations in particular subsets of the pixels.

Most known works in image modelling (see, for instance, compehensive surveys of DUBES and JAIN [1989], TUCERYAN and JAIN [1993], LI [1995], and WINKLER [1995]) exploit two types of the Markov/Gibbs models with pairwise pixel interactions, namely, the autobinomial and the autonormal (Gauss Markov) models, called *automodels* by BESAG ([1974]). Restriction to the models with pairwise interactions is common, not only to

the image processing domain, because of lesser theoretical and computational complexity (AVERINTSEV [1970, 1972], DOBRUSHIN [1968], DOBRUSHIN and PIGOROV [1975], CHELLAPPA and JAIN [1993]). But, the automodels, borrowed from statistical physics and measurement theory, ignore that the pixel interactions have no physical meaning and cannot be described adequately by the "physics-based" potentials based on signal products. In physics, a direct dependency between the interaction strength and the signal product is justified. But, it is obvious that the product of gray levels does not relate to their frequencies in the image. Also, the interaction structure of the automodels is predefined to a considerable extent. The autobinomial model exploits only two or four clique families that form, respectively, the nearest 4- or 8-neighborhood graph. Thus, in principle, it cannot take account of more distant neighbors typical to most natural textures. The autonormal, or Gauss-Markov model is much more flexible because it assumes that all the pixels within a given rectangular, square, or circular window form a clique, that is, are the neighbors of each other. But, there are no appropriate ways to choose the most characteristic window for a given texture type (except of a direct exhaustion of all possible variants that is computationally inappropriate) or to exclude less significant interactions from the window. Also, the automodels meet computational difficulties in learning the potentials (say, in solving large systems of non-linear equations to find the MLE of parameters for the Gauss-Markov model).

Due to uncapability to learn the interaction structure, these traditional models do not provide a uniform framework for describing different regular and irregular image patterns and, therefore, have rather limited possibilities in modelling the natural textures.

2.2 Models with multiple pairwise interactions

Regarding the homogeneous textures, these models overcome some drawbacks of the traditional automodels in two directions (GIMEL'FARB [1996a, 1996b]): (*i*) they allow to learn both the characteristic interaction structure and the Gibbs potentials from a given training sample, and (*ii*) some simple gray range transformations (such as arbitrary shifts or linear stretches of the image gray ranges) that have no effect on the texture type can be taken into account by choosing a particular form of the potentials or by involving a particular non-Markov Gibbs model. The Gray Level Co-occurrence Histograms (GLCHs) or the GLDHs are *sufficient statistics* for them. Therefore, it is these models that appear implicitly in well-known approaches to describe the grayscale textures by these histograms (see, for instance, HARALICK [1979]). Below, we overview two such models: the Markov/Gibbs one that admits the gray range shifts and the non-Markov model admitting the gray range stretches.

2.2.1 Markov/Gibbs model of grayscale images

Generally, the admissible gray range shifts result in a non-Markov image model. But, the Markov/Gibbs case can be preserved still by assuming that the interaction strength is invariant with respect to the shifts. Under this additional assumption, the potentials for the pixelwise interactions are set to zero and potentials for the pairwise interactions depend only on the gray level differences. The corresponding Markov/Gibbs model is as follows (GIMEL'FARB [1996a]):

$$\Pr(\mathbf{g}|\mathbf{V}) \quad = \quad \frac{1}{Z_{\mathbf{V}}} \cdot \exp\left(\sum_{a\in\mathbf{A}} e_a(\mathbf{g}|\mathbf{V})\right). \tag{1}$$

Here, $\mathbf{V} = (V_a(d) : \ d \in \mathbf{D};\ a \in \mathbf{A})$ is the vector of the centered potential values for the gray level differences in the cliques of all the given pairwise clique families $\mathbf{C}$ (the potential centering in this and subsequent models is discussed in more detail in Section 2.3 below), $e_a(\mathbf{g}|\mathbf{V}) = \sum_{(i,j)\in\mathbf{C}_a} V_a(g(i) - g(j))$ is a partial (Gibbs) *energy* of the pairwise pixel interactions, and $Z_{\mathbf{V}}$ is a normalizing factor, also called partition function in statistical physics (ISIHARA [1971], BESAG[1974]). The total Gibbs energy of the interactions in (1) is as follows: $E(\mathbf{g}|\mathbf{V}) = \sum_{a\in\mathbf{A}} e_a(\mathbf{g}|\mathbf{V})$.

Both the partial and total Gibbs energies can be represented by dot products of the potential vector and the histogram vectors for the images:

$$\forall_{a\in\mathbf{A}} \ \ e_a(\mathbf{g}|\mathbf{V}) = \mathbf{V} \bullet \mathbf{H}_a(\mathbf{g}) = \sum_{d\in\mathbf{D}} V_a(d) \cdot H_a(d|\mathbf{g}) \equiv \sum_{d\in\mathbf{D}} V_a(d) \cdot H_{a,\mathrm{cn}}(d|\mathbf{g})$$

so that the total Gibbs energy is as follows:

$$E(\mathbf{g}|\mathbf{V}) \ = \ \sum_{a\in\mathbf{A}} e_a(\mathbf{g}|\mathbf{V}) \ = \ \mathbf{V} \bullet \mathbf{H}(\mathbf{g}) \ \equiv \ \mathbf{V} \bullet \mathbf{H}_{\mathrm{cn}}(\mathbf{g}).$$

Here, the subscript "cn" indicates the centering. Therefore, the model (1) can be written in the following equivalent form of the exponential family distribution (these families were studied in detail by BARNDORFF-NIELSEN [1978]):

$$\Pr(\mathbf{g}|\mathbf{V}) \quad = \quad \frac{1}{Z_{\mathbf{V}}} \cdot \exp\left(\mathbf{V} \bullet \mathbf{H}_{\mathrm{cn}}(\mathbf{g})\right). \tag{2}$$

where $\mathbf{H}(\mathbf{g}) = \{\mathbf{H}_a(\mathbf{g}) : \ a \in \mathbf{A}\}$ denotes the vector of the (centered) GLDHs for the image $\mathbf{g}$. This representation shows that the (centered) GLDHs, collected over the clique families, $\mathbf{C}$ form the sufficient statistic for the model.

2.2.2 Non-Markov Gibbs model of grayscale images

Under the admissible changes of the image gray ranges, each image $\mathbf{g}$ has to be equivalent, by its Gibbs probability, to a particular reference sample $\mathbf{g}^{\mathrm{rf}}$ obtained by normalizing the initial gray range $[q_{\min}(\mathbf{g}), q_{\max}(\mathbf{g})]$ into a given reference range, say, $[0, q_{\max}]$:

$$g^{\mathrm{rf}}(i) = \frac{g(i) - q_{\min}(\mathbf{g})}{q_{\max}(\mathbf{g}) - q_{\min}(\mathbf{g})} \cdot q_{\max}.$$

The non-Markov Gibbs models that allow for these gray range changes are obtained by embedding the normalization $\mathbf{g} \rightarrow \mathbf{g}^{\mathrm{rf}}$ directly into the Gibbs potentials (GIMEL'FARB [1996b]). After the normalization, each pixel depends on all other pixels in such a way that the local interactions are supplemented with a lattice-wide interaction for getting the minimum $q_{\min}(\mathbf{g})$ and maximum $q_{\max}(\mathbf{g})$ gray levels. But, the latter interaction manifests itself only if the pixel supports the solitary minimum or maximum gray level in the lattice. Otherwise, only the local interactions have to be taken into account.

This approach leads to the following non-Markov Gibbs model of the grayscale images (GIMEL'FARB [1996b]):

$$\Pr(\mathbf{g}|\mathbf{V}) = \frac{1}{Z_{\mathbf{V}}} \cdot \exp\left(e(\mathbf{g}^{\mathrm{rf}}|\mathbf{V}) + \sum_{a \in \mathbf{A}} e_a(\mathbf{g}^{\mathrm{rf}}|\mathbf{V}) \right) \tag{3}$$

where $e(\mathbf{g}^{\mathrm{rf}}|\mathbf{V}) = \sum_{i \in \mathbf{R}} V(g^{\mathrm{rf}}(i))$ and $e_a(\mathbf{g}^{\mathrm{rf}}|\mathbf{V}) = \sum_{(i,j) \in \mathbf{C}_a} V_a(g^{\mathrm{rf}}(i), g^{\mathrm{rf}}(j))$ denote, respectively, the partial energy of the pixelwise interactions and of the pairwise pixel interactions for the clique family $\mathbf{C}_a$. Here, $\mathbf{V} = \left(V(q) : q \in \mathbf{G};\ V_a(q, q') : (q, q') \in \mathbf{G}^2;\ a \in \mathbf{A}\right)$ is the vector of the centered potential values for the gray levels and gray level co-occurrences in the cliques. The exponential family representation, similar to (2), shows that the GLH and a subset of the GLCHs for the reference grayscale image $\mathbf{g}^{\mathrm{rf}}$ form the sufficient statistic for this model.

The model (3) is simplified by an additional assumption that the potentials for the pixel pairs depend only on the signal differences $d = g^{\mathrm{rf}}(i) - g^{\mathrm{rf}}(j)$ so that the potential vector $\mathbf{V} = (V(q) : q \in \mathbf{G};\ V_a(d) : d \in \mathbf{D};\ a \in \mathbf{A})$ contains the centered potential values for the gray levels and gray level differences in the cliques. The partial energies for the clique families are as follows:

$$e_a(\mathbf{g}^{\mathrm{rf}}|\mathbf{V}) = \sum_{(i,j) \in \mathbf{C}_a} V_a(g^{\mathrm{rf}}(i) - g^{\mathrm{rf}}(j)).$$

The simplified non-Markov Gibbs model is represented by the equivalent exponential family distribution:

$$\Pr(\mathbf{g}|\mathbf{V}) = \frac{1}{Z_{\mathbf{V}}} \cdot \exp\left(\mathbf{V} \bullet \mathbf{H}_{\rm cn}(\mathbf{g})\right) \tag{4}$$

where $\mathbf{H}_{\rm cn}(\mathbf{g}) = \{H_{\rm cn,pix}(q|\mathbf{g}) : \ q \in \mathbf{Q};\ H_{{\rm cn},a}(d|\mathbf{g}) : \ d \in \mathbf{D};\ a \in \mathbf{A}\}$ is a vector of the centered GLH and GLDHs for the reference image $\mathbf{g}$. Thus, these centered GLH and GLDHs form the *sufficient statistic* for the model (3). Only this simplified non-Markov model will be considered below.

The basic features of the non-Markov model (3) are similar or obtained with minor changes from the like features of the Markov/Gibbs one (1).The major distinction between them is in the stochastic relaxation techniques for generating the samples under a given GPD. Each relaxation step involves, in the Markov/Gibbs case, a summation of the potentials only over a local neighbourhood of the current pixel. The neighbourhood is formed by the cliques containing this pixel. In the non-Markov case, the local summation holds for all the pixels of the reference training sample, except for the pixels of the solitary maximum or minimum signal. Only in this (and rather rare) case, the actual neighbourhood of the pixel is lattice-wide so that the potentials are summed up over the total lattice. Thus, the computational complexity of the relaxation is $O(|\mathbf{R}|)$ in both the cases and does not increase substantially in the non-Markov case as compared to the Markov/Gibbs one.

2.3 Potential centering

Generally, if the potentials are unconstrained then the GPDs (1) – (4) give non-unique representations of the GRFs. In other words, the same GRF is represented by the different potentials $\mathbf{V}$. For instance, an arbitrary constant, added to all the potential values for a clique family $\mathbf{C}_a$, does not change the image probability because it is reduced from the nominator and denominator of the GPDs.

Usually, the unique representation is obtained by using relative Hamiltonians, that is, relative total Gibbs energies $E_{\rm rel}(\mathbf{g}|\mathbf{V})$, as the exponents (see, for instance, DOBRUSHIN and PIGOROV [1975]). One sample $\mathbf{g}_{\rm bas} \in \mathbf{G}$ is chosen as a special base to be reduced from the nominator and denominator of the GPD so that the exponents are given by the differences $E_{\rm rel}(\mathbf{g}) = E(\mathbf{g}) - E(\mathbf{g}_{\rm bas})$. It is easily seen that the relative energies are represented equivalently as follows:

$$E_{\rm rel}(\mathbf{g}) = \mathbf{V} \bullet (\mathbf{H}(\mathbf{g}) - \mathbf{H}(\mathbf{g}_{\rm bas}))$$

.

The latter representation shows that the relative energies in the models (1) – (4) are invariant to the following *potential centering* introduced in (GIMEL'FARB [1996a, 1996b]):

$$\sum_{q\in\mathbf{Q}} V(q) = 0;\ \forall_{a\in\mathbf{A}} \sum_{q,q'\in\mathbf{Q}^2} V_a(q,q') = \sum_{d\in\mathbf{D}} V_a(d) = 0 \tag{5}$$

as well as to the like centering of the initial GLH and GLDHs, too. Both the centerings are due to the obvious relations:

$$\sum_{q\in\mathbf{Q}} H(q|\mathbf{g}) = |\mathbf{R}|;\ \forall_{a\in\mathbf{A}} \sum_{q,q'\in\mathbf{Q}^2} H_a(q,q'|\mathbf{g}) = \sum_{d\in\mathbf{D}} H_a(d|\mathbf{g}) = |\mathbf{C}_a|$$

for the initial histograms. The centered vectors $\mathbf{V}$ and $\mathbf{H}_{\rm cn}(\mathbf{g})$ in (2) and (4) lie in the same G-dimensional vector subspace $\mathcal{S} \subset \mathcal{R}^{G+|\mathbf{A}|+1}$ where $G = q_{\max}\cdot(2\cdot|\mathbf{A}|+1)$ and $\mathcal{R}$ denotes real numbers.

Notice that the centering conditions (5) are the only constraints on the potentials imposed by the relative Hamiltonians. As it will be shown below, in Section 4, we not necessarily need the unique representation of the GPD for simulating the samples by a pixelwise stochastic relaxation or for estimating the model parameters. The former case involves the differences between the total energies and the latter one exploits the differences between the histograms so that the "base" energy $E(\mathbf{g}_{\rm bas})$ or the "base" histogram $\mathbf{H}(\mathbf{g}_{\rm bas})$ are excluded from the consideration. But, of course, the centering conditions (5) on the potentials are still retained and have to be taken into account.

2.3.1 Markov/Gibbs model of region maps

The region maps differ from the grayscale images only in a physical meaning of the signals: the gray levels, measured in a metric scale, represent brightness of the sensed objects over an image sensor Field-Of-View (FOV) whereas the region labels, measured in a nominal scale, indicate parts of the FOV occupied by different objects (in our case, for instance, by different homogeneous textures).

The pairwise label interactions depend, generally, on label co-occurrences. For simplicity, let us assume that the pixelwise interactions are the same for all the regions (so that the corresponding potentials can be set to zero) and that the inter-region interactions depend only on label coincidences (that is, for each the region $k \in \mathbf{K}$, all the "foreign" regions

are equivalent). Under these additional assumptions, the Markov/Gibbs model of the region maps is as follows:

$$\Pr(\mathbf{l}|\mathbf{V}) \;=\; \frac{1}{Z_{\mathbf{V}}}\cdot\exp\left(E(\mathbf{l}|\mathbf{V})\right) \;\equiv\; \frac{1}{Z_{\mathbf{V}}}\cdot\exp\left(\sum_{a\in\mathbf{A}} e_a(\mathbf{l}|\mathbf{V})\right). \tag{6}$$

Here, $\mathbf{V} = (V_{a,\alpha}(k):\ k\in\mathbf{K};\ \alpha\in\{0,1\};\ a\in\mathbf{A})$ is the vector of the centered potential values for the label coincidences $\alpha=\delta(l(i)-l(j))$; $k=l(i)$, in the cliques $(i,j)\in\mathbf{C}_a\subset\mathbf{C}$ and $e_a(\mathbf{l}|\mathbf{V}) = \sum\limits_{(i,j)\in\mathbf{C}_a} V_{a,\delta(l(i)-l(j)}(l(i))$ is the partial energy of the pairwise pixel interactions. The total interaction energy in (6) is as follows: $E(\mathbf{l}|\mathbf{V}) = \sum\limits_{a\in\mathbf{A}} e_a(\mathbf{l}|\mathbf{V})$.

The model (6) has the following exponential family representation:

$$\Pr(\mathbf{l}|\mathbf{V}) \;=\; \frac{1}{Z_{\mathbf{V}}}\cdot\exp\left(\mathbf{V}\bullet\mathbf{H}_{\mathrm{cn}}(\mathbf{l})\right)). \tag{7}$$

where $\mathbf{H}_{\mathrm{cn}}(\mathbf{l}) = \{\mathbf{H}_{\mathrm{cn},a}(\mathbf{l}):\ a\in\mathbf{A}\}$ denotes the vector of the (centered) RLCHs for the map $\mathbf{l}$:

$$\mathbf{H}_{\mathrm{cn},a}(\mathbf{l}) = \{H_{a,\alpha}(k|\mathbf{l}):\ k\in\mathbf{K};\ \alpha\in\{0,1\}\}$$

where

$$H_{a,\alpha}(k|\mathbf{l}) = \sum_{(i,j)\in\mathbf{C}_a}\delta(k-l(i))\cdot\delta(\alpha-\delta(l(i)-l(j))).$$

This representation shows that the (centered) RLCHs collected over the clique families $\mathbf{C}$ form the sufficient statistic for the model.

The potential centering in this model is as follows:

$$\forall_{a\in\mathbf{A}}\ \sum_{k\in\mathbf{K}}\sum_{\alpha\in\{0,1\}} V_{a,\alpha}(k) = 0. \tag{8}$$

One can simplify the model (6) by assuming the labels from the different regions are mutually independent. In this case the potentials for the inter-region interactions are set to zero so that only the intra-region potentials in 8 have to be centered:

$$\forall_{a\in\mathbf{A}}\ \ \forall_{k\in\mathbf{K}}\ \ V_{a,0}(k) = 0;\ \ \sum_{k\in\mathbf{K}} V_{a,1}(k) = 0. \tag{9}$$

3 Piecewise-Homogeneous Gibbs Fields

The above Gibbs models of spatially homogeneous stochastic textures are easily extended onto more complex piecewise-homogeneous textures (GIMEL'FARB [1996c]). The piecewise homogeneous grayscale texture can be represented by a superimposed pair containing the grayscale image and the supporting region map. The map shows the homogeneous regions in the lattice. In this case, there is a signal pair per pixel $i \in \mathbf{R}$ containing the gray level $q = g(i)$ for the grayscale image $\mathbf{g}$ and the region label $k = l(i)$ for the region map l. Therefore, the pixelwise interaction involves this signal pair $(q = g(i),\ k = l(i))$ and every pairwise interaction deals with the signal quadruple $(q = g(i),\ q' = g(j),\ k = l(i),\ k' = l(j))$ where $(i, j) \in \mathbf{C}_a$.

3.1 Joint model of grayscale images and region maps

The joint GPD for describing the piecewise-homogeneous texture $\mathbf{g}$ and its corresponding region map l is easily obtained by generalizing the GPDs for the homogeneous textures of Section 2. Each pair $(\mathbf{g}, \mathrm{l})$ is considered as a sample of the GRF with multiple pairwise interactions. The parent population is formed by the Cartesian product $\mathbf{G} \times \mathbf{L}$.

This model is simplified in (GIMEL'FARB [1996c]) by additional assumptions that the pairwise pixel interactions depend only (*i*) on the coincidence of the region labels and (*ii*) on the gray level differences. In other words, only two types $\alpha \in \{0, 1\}$ of the label interactions: the intra-region interaction ($\alpha = \delta(l(i) - l(j)) = 1$, or $l(i) = l(j) = k$) and the inter-region interaction ($\alpha = \delta(l(i) - l(j)) = 0$, or $l(i) = k \neq l(j)$) are taken into account for each region k. Then in all, only $(q_{\max} + 1) \cdot (k_{\max} + 1) + 2 \cdot (2q_{\max} + 1) \cdot (k_{\max} + 1)$ potential values have to be learnt per clique family instead of $(q_{\max} + 1) \cdot (k_{\max} + 1) + (q_{\max} + 1)^2 \cdot (k_{\max} + 1)^2$ for the general case of the label and gray level co-occurrences.

Let $V(q, k)$ and $V_a(q, q', k, k') \equiv V_{a,\alpha=\delta(k-k')}(d = q - q', k)$ be the potentials for the signal pair (q, k) and the quadruple (q, q', k, k'), respectively. The resulting joint GPD is as follows:

$$\Pr(\mathbf{g}, \mathrm{l}|\mathbf{V}) = \frac{1}{Z_{\mathbf{V}}} \cdot \exp\left(e(\mathbf{g}, \mathrm{l}|\mathbf{V}) + \sum_{a \in \mathbf{A}} e_a(\mathbf{g}, \mathrm{l}|\mathbf{V})\right) \tag{10}$$

where

$$e(\mathbf{g}, \mathrm{l}|\mathbf{V}) = \sum_{i \in \mathbf{R}} V(g(i), l(i))$$

and

$$e_a(\mathbf{g},\mathrm{l}|\mathbf{V}) = \sum_{(i,j)\in\mathbf{C}_a} V_{a,\delta(l(i)-l(j))}(g(i)-g(j),l(i))$$

are the partial energies for the pixelwise interactions and for the pairwise pixel interactions in the clique family $\mathbf{C}_a$, respectively, and $\mathbf{V} = \{V(q,k):\ (q,k)\in\mathbf{Q}\times\mathbf{K};\ V_{a,\alpha}(d,k):\ a\in\mathbf{A};\ \alpha\in\{0,1\};\ (d,k)\in\mathbf{D}\times\mathbf{K}\}$ is the vector of the centered potentials. For brevity sake, here and below we omit the superscript "rf" for the reference samples of the grayscale images.

This model is represented by the following exponential family distribution:

$$\Pr(\mathbf{g},\mathrm{l}|\mathbf{V}) = \frac{1}{Z_{\mathbf{V}}}\cdot\exp\left(\mathbf{V}\bullet\mathbf{H}_{\mathrm{cn}}(\mathbf{g},\mathrm{l})\right) \tag{11}$$

where $\mathbf{H}_{\mathrm{cn}}(\mathbf{g},\mathrm{l}) = \{\mathbf{H}_{\mathrm{cn,pix}}(\mathbf{g},\mathrm{l});\ \mathbf{H}_{\mathrm{cn},a}(\mathbf{g},\mathrm{l}):\ a\in\mathbf{A}\}$ is the vector of the centered joint GL/RLH and GLD/RLCHs. These histograms are the sufficient statistic for the model. The potential centering here is quite similar to the centering in (5):

$$\sum_{k\in\mathbf{K}}\sum_{q\in\mathbf{Q}} V(q,k) = 0;\ \ \forall_{a\in\mathbf{A}}\sum_{k\in\mathbf{K}}\sum_{\alpha=0}^{1}\sum_{d\in\mathbf{D}} V_{a,\alpha}(d,k) = 0. \tag{12}$$

For simplicity, the characteristic interaction structure in this model is assumed to be the same in all the regions. But, it is not difficult to extend the model so that each region has its own interaction structure, that is, a distinct subset $\mathbf{A}_k$ of the clique families.

3.2 Conditional models of grayscale images and region maps

The joint Gibbs model (10) is easily reduced to conditional models by fixing either the grayscale image $\mathbf{g} = \mathbf{g}^\circ$ or the region map $\mathrm{l} = \mathrm{l}^\circ$. For brevity sake, only their exponential family representations are presented below. The conditional model of the images, given a region map l°, is as follows:

$$\Pr(\mathbf{g}|\mathbf{V},\mathrm{l}^\circ) = \frac{1}{Z_{\mathbf{V},\mathrm{l}^\circ}}\cdot\exp\left(\mathbf{V}\bullet\mathbf{H}_{\mathrm{cn}}(\mathbf{g},\mathrm{l}^\circ)\right). \tag{13}$$

The model of the region maps, given a reference image $\mathbf{g}^\circ$, has a symmetric form:

$$\frac{1}{Z_{\mathbf{V},\mathbf{g}^\circ}}\cdot\Pr(\mathrm{l}|\mathbf{V},\mathbf{g}^\circ) = \exp\left(\mathbf{V}\bullet\mathbf{H}_{\mathrm{cn}}(\mathbf{g}^\circ,\mathrm{l})\right). \tag{14}$$

Both these models differ from the joint model (11) only in the parent populations ($\mathbf{G}$ and $\mathbf{L}$, respectively), in the partition functions, and in the potential centering.

For the model (13), the homogeneous regions are fixed for all the samples $\mathbf{g} \in \mathbf{G}$. The clique family $\mathbf{C}_a$ is partitioned onto $|\mathbf{K}| = k_{\max} + 1$ fixed subfamilies $\mathbf{C}_{a,k^\circ} = \{(i,j) : (i,j) \in \mathbf{C}_a;\ l^\circ(i) = k^\circ\}$, each containing the cliques from a single region $k^\circ \in \mathbf{K}$. Therefore, the potentials are individually centered for each the subfamily as follows:

$$\forall_{k^\circ \in \mathbf{K};\, a \in \mathbf{A}} \quad \sum_{d \in \mathbf{D}} \sum_{\alpha=0}^{1} V_{a,\alpha}(d, k^\circ) = 0. \tag{15}$$

Generally, each region $k^\circ \in \mathbf{K}$ may have its own interaction structure $\mathbf{A}_{k^\circ}$ so that the model (13) possesses the following form:

$$\begin{aligned} \Pr(\mathbf{g}|\mathbf{V}, \mathrm{l}^\circ) \quad = \quad & \frac{1}{Z_{\mathbf{V},\mathrm{l}^\circ}} \cdot \exp\left(\sum_{i \in \mathbf{R}} (V(g(i), l^\circ(i)) + \right. \\ & \left. \sum_{k \in \mathbf{K}} \delta(k - l^\circ(i)) \sum_{a \in \mathbf{A}_k} \sum_{j:(i,j) \in \mathbf{C}_a} V_{a,\delta(k-l^\circ(j))}(g(i) - g(j), k) \right) \right) \end{aligned} \tag{16}$$

Let us assume, for simplicity, that the gray levels from the different regions are mutually independent. This allows for setting to zero the potentials for all the inter-region interactions: $\forall_{k^\circ \in \mathbf{K};\, a \in \mathbf{A};\, d \in \mathbf{D}}\ V_{a,\alpha}(d, k^\circ) = 0$ and for learning the interaction structure and potentials independently in each homogeneous region of the training pair $(\mathbf{g}^\circ, \mathrm{l}^\circ)$. In this case, the piecewise-homogeneous image with the known region map can be modeled by adapting first the model (3) to each type of the homogeneous textures and then by using the learnt interaction structure(s) and Gibbs potentials in the conditional model (16).

The model (14) is quite symmetric to the model (13) in that each clique family is partitioned onto $|\mathbf{D}| = 2q_{\max} + 1$ fixed subfamilies. In this case, the subfamily $\mathbf{C}_{a,d^\circ} = \{(i,j) : (i,j) \in \mathbf{C}_a;\ \ g(i) - g(j) = d^\circ\}$, $d^\circ \in \mathbf{D}$, contains the cliques with the constant signal difference in the given grayscale image $\mathbf{g}^\circ$. The potential centering is as follows:

$$\forall_{d^\circ \in \mathbf{D};\, a \in \mathbf{A}} \quad \sum_{k \in \mathbf{K}} \sum_{\alpha=0}^{1} V_{a,\alpha}(d^\circ, k) = 0. \tag{17}$$

In this model, the characteristic interaction structure may generally depend on the signal differences d° so that the Gibbs energy for the pairwise interactions is computed over a

union $\mathbf{A} = \bigcup_{d^\circ \in \mathbf{D}} \mathbf{A}_{d^\circ}$ of all these structures. Each clique $(i,j) \in \mathbf{A}_{d^\circ}$ is taken into account if and only if $g^\circ(i) - g^\circ(j) = d^\circ$. In this case, the extended conditional GPD of the region maps takes the following form:

$$\Pr(\mathbf{l}|\mathbf{V},\mathbf{g}^\circ) = \frac{1}{Z_{\mathbf{V},\mathbf{g}^\circ}} \cdot \exp\left(\sum_{i\in\mathbf{R}} V(g^\circ(i),l(i)) + \right.$$

$$\left. \sum_{d\in\mathbf{D}} \sum_{a\in\mathbf{A}_d} \sum_{(i,j)\in\mathbf{C}_a} V_{a,\delta(l(i)-l(j))}(d,l(i)) \cdot \delta\left(d-(g^\circ(i)-g^\circ(j))\right)\right) \tag{18}$$

The simplifying assumption that the different region labels are mutually independent allows for setting to zero the potentials of the inter-region interactions: $\forall_{d^\circ\in\mathbf{D};\, a\in\mathbf{A};\, k\in\mathbf{K}}$ $V_{a,0}(d^\circ,k) = 0$ so that the intra-region potentials possess the following centering:

$$\forall_{d^\circ\in\mathbf{D};\, a\in\mathbf{A}} \quad \sum_{k\in\mathbf{K}} V_{a,1}(d^\circ,k) = 0. \tag{19}$$

4 Learning the Model Parameters

All the above models possess almost the same learning procedures which allow for recovering both the interaction structure and strengths from given training samples (a grayscale image for the homogeneous texture or a pair of an image and corresponding region map for the piecewise-homogeneous one). This procedure, proposed for the homogeneous textures by GIMEL'FARB [1996a], starts from an analytic first approximation of the maxumum likelihood estimate (MLE) of the potentials. This approximation, computed from the GLDHs for the model (1) and from the GLH and GLDHs for the model (3), enables to compare relative strengths of a great many possible pairwise interactions and recover most characteristic ones to represent a given texture type. Then, the desired MLE of the potentials for the chosen families $\mathbf{C}_a$ is refined by a stochastic approximation technique similar to the one introduced by YOUNES [1988].

As shown by BARNDORFF-NIELSEN ([1978]), the GPD in (4) is the regular exponential family distribution with minimal canonical parameter $\mathbf{V}$ and minimal sufficient statistic $\mathbf{H}_{\mathrm{cn}}(\mathbf{g})$ if and only if the following conditions are both satisfied: (*i*) the vectors $\mathbf{V}$ are affinely independent and (*ii*) the vectors $\mathbf{H}_{\mathrm{cn}}(\mathbf{g})$ are affinely independent. These conditions ensure that the distribution is strictly log-concave with respect to the centered potential vectors $\mathbf{V}$. The affine independence of the potential vectors $\mathbf{V}$ holds because there are no

restrictions on them except for the centering (5). The log-concavity of the above GPDs is obvious because the Hessian for each corresponding log-likelihood function is equal to the covariance matrix for the histogram vectors, which is taken with the negative sign, and therefore is always negative semi-definite. But, under the second condition, it is negative definite. Therefore, only the affine independence of the centered histogram vectors $\mathbf{H}_{\mathrm{cn}}(\mathbf{g})$ for the parent population $\mathbf{G}$ has to be proven. Below, we prove this for the models (1), (3), and (6). The proof for the first two models has also been given in (GIMEL'FARB, SCHMIDT, and BRAUNMANDL [1997]).

4.1 Affine independence of the potentials and histograms

Lemma 1 *Let the lattice* $\mathbf{R}$ *contain, at least,* $2 \cdot q_{\max}$ *cliques of each family and allow to arrange them into separate pairs sharing each the same pixel. Then the histogram vectors* $\mathbf{H}_{\mathrm{cn}}(\mathbf{g})$*;* $\mathbf{g} \in \mathbf{G}$*, are affinely independent in the subspace* $\mathcal{S} \subset \mathcal{R}^{G+|\mathbf{A}|+1}$*;* $\dim(\mathcal{S}) = G$*, of the centered vectors where* $G = q_{\max} \cdot (2 \cdot |\mathbf{A}| + 1)$.

For proving Lemma 1, let us form in the subspace $\mathcal{S}$ an orthogonal basis with G vectors stratified into $1 + |\mathbf{A}|$ groups. The first group, corresponding to the GLH-part of the histogram vectors, possesses $q_{\max}$ vectors. The other groups correspond to the GLDH-parts for the different clique families and have each $2 \cdot q_{\max}$ vectors. The group $a + 1$ corresponds to the clique family $\mathbf{C}_a$. Each basis vector is formed by concatenating one subvector of length $1 + q_{\max}$ for the GLH-part of the histogram vectors and $|\mathbf{A}|$ subvectors of length $1 + 2 \cdot q_{\max}$ for their GLDH-parts. Below, we consider only the GLDH-parts in the histogram vectors because the proof for the GLH-part is quite similar.

In the basis vectors from the first group, all the subvectors, except for the first one, are zero-valued. In the basis vectors from the group $a + 1$, all the subvectors, except for the subvector $a + 1$, are also zero-valued. Possible nontrivial $2 \cdot q_{\max}$ subvectors $\{\mathbf{b}_{a,q}, \mathbf{c}_{a,q} : q = 1, \ldots, q_{\max}\}$ that form the orthogonal (sub)basis in $\mathcal{S}$ are shown in Table 1.

It is evident that the difference histogram vectors $\mathbf{\Delta}(\mathbf{g}, \mathbf{g}') = \mathbf{H}_{\mathrm{cn}}(\mathbf{g}) - \mathbf{H}_{\mathrm{cn}}(\mathbf{g}')$ for the pairs of the reference image samples lie in the subspace $\mathcal{S}$, too. To prove the affine independence of the centered histogram vectors, it is sufficient to show that this holds for their difference vectors. In other words, it is sufficient to show that all the basis subvectors $\{\mathbf{b}_{a,q},\ \mathbf{c}_{a,q} :\ q = 1, \ldots, q_{\max}\}$ from Table 1 appear in the difference vectors for each clique family $\mathbf{C}_a$.

The image samples giving the desired difference vectors are formed as follows.

Table 1: Basis subvectors to prove Lemma 1 (for brevity, $q_{\max}$ is replaced by ψ)

GLD d	$-\psi$	$-\psi+1$	$-\psi+2$	...	-1	0	1	...	$\psi-2$	$\psi-1$	ψ
$\mathbf{b}_{a,1}$	-1	0	0	...	0	0	0	...	0	0	1
$\mathbf{b}_{a,2}$	0	-1	0	...	0	0	0	...	0	1	0
$\mathbf{b}_{a,3}$	0	0	-1	...	0	0	0	...	1	0	0
...	...	...	...	...	...	...	...	...	...	...	...
$\mathbf{b}_{a,\psi}$	0	0	0	...	-1	0	1	...	0	0	0
$\mathbf{c}_{a,1}$	-1	1	0	...	0	0	0	...	0	1	-1
$\mathbf{c}_{a,2}$	-1	-1	2	...	0	0	0	...	2	-1	-1
...	...	...	...	...	...	...	...	...	...	...	...
$\mathbf{c}_{a,\psi-1}$	-1	-1	-1	...	$\psi-1$	0	$\psi-1$	...	-1	-1	-1
$\mathbf{c}_{a,\psi}$	-1	-1	-1	...	-1	2ψ	-1	...	-1	-1	-1

- (*i*) Let the reference image sample $\mathbf{g}$ have two contiguous regions with the constant signals q° and $q^\circ+q$ in the pixels, respectively, with the exception of two pixels with the maximum ($q_{\max}$) and minimum (0) values within the region with the signal q°, and let the sample $\mathbf{g}'$ have just the same regions but with the signals q° and $q^\circ-q$. The value q° is chosen so that all three values $q^\circ-q$, q°, and $q^\circ+q$ are in the set $\mathbf{Q}$. Then, for $q=1,\ldots,q_{\max}$ and these sample pairs, the difference subvectors, to within a certain scaling factor, are the same as the basis subvectors $\{\mathbf{b}_{a,q}:\ q=1,\ldots,q_{\max}\}$.

- (*ii*) Let the reference sample $\mathbf{g}$ contain all zero–valued signals, except for q pixels with the same signal $q_{\max}-q$ and one pixel with the signal $q_{\max}$. The q pixels are arranged in such a way that each pixel belongs to its "own" pair of the cliques from the family $\mathbf{C}_a$ having the signal configurations $(0, q_{\max}-q)$ and $(q_{\max}-q, 0)$, respectively. The sample $\mathbf{g}'$ has the same form but the signals in the above q pixels possess the successive values $q_{\max},\ldots,q_{\max}-q+1$. The difference subvectors, for $q=1,\ldots,q_{\max}$, are the same as the basis subvectors $\{\mathbf{c}_{a,q}:\ q=1,\ldots,q_{\max}\}$.

Therefore, all the basis subvectors from Table 1, for any second–order clique family, take part in the non–zero difference vectors $\mathbf{\Delta}(\ldots)$ and, hence, the GLDH-parts of the centered histogram vectors are affinely independent in the vector subspace $\mathcal{S}$. The independence of the GLH-part is proven in a similar way. This proof can be adapted also to the conditional models (13) and (16) of the grayscale images under a given region map.

The like Lemma can be formulated for the region map model (6). In this case, the orthogonal basis in the subspace $\mathcal{S}$ contains $L = (2 \cdot k_{\max} + 1) \cdot |\mathbf{A}|$ centered basis vectors.

Lemma 2 *Let the lattice* $\mathbf{R}$ *contains, at least,* $3 \cdot k_{\max}$ *cliques of each family and allow to arrange them into separate pairs sharing each the same pixels. Then the histogram vectors* $\mathbf{H}_{\mathrm{cn}}(\mathrm{l})$*;* $\mathrm{l} \in \mathbf{L}$*, are affinely independent in the subspace* $\mathcal{S} \subset \mathcal{R}^{L+|\mathbf{A}|}$*;* $\dim(\mathcal{S}) = L$*, of the centered vectors.*

For proving Lemma 2, let us stratify the basis vectors into $|\mathbf{A}|$ groups (one group per clique family) having each $2 \cdot k_{\max} + 1$ vectors. Each basis vector is formed by concatenation of $|\mathbf{A}|$ subvectors of length $2 \cdot (k_{\max} + 1)$. All these subvectors are zero-valued except for the subvector a in the basis vectors from the group a that corresponds to the clique family $\mathbf{C}_a$. Table 2 contains possible $2 \cdot k_{\max} + 1$ subvectors $\{\mathbf{b}_{a,k} : k = 0, \ldots, k_{\max};\ \mathbf{c}_{a,k} : k = 1, \ldots, k_{\max}\}$ giving the desired orthogonal (sub)basis.

Table 2: Basis subvectors to prove Lemma 2 (for brevity, $k_{\max}$ is replaced by ϕ)

Region	0	1	2	…	$\phi-1$	ϕ	0	1	2	…	$\phi-1$	ϕ
α	1						0					
$\mathbf{b}_{a,0}$	1	0	0	…	0	0	-1	0	0	…	0	0
$\mathbf{b}_{a,1}$	0	1	0	…	0	0	0	-1	0	…	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…
$\mathbf{b}_{a,\phi-1}$	0	0	0	…	1	0	0	0	0	…	-1	0
$\mathbf{b}_{a,\phi}$	0	0	0	…	0	1	0	0	0	…	0	-1
$\mathbf{c}_{a,1}$	-1	1	0	…	0	0	-1	1	0	…	0	0
$\mathbf{c}_{a,2}$	-1	-1	2	…	0	0	-1	-1	2	…	0	0
…	…	…	…	…	…	…	…	…	…	…	…	…
$\mathbf{c}_{a,\phi-1}$	-1	-1	-1	…	$\phi-1$	0	-1	-1	-1	…	$\phi-1$	0
$\mathbf{c}_{a,\phi}$	-1	-1	-1	…	-1	ϕ	-1	-1	-1	…	-1	ϕ

In this case, we also show that all the basis subvectors from Table 2 appear in the difference histogram vectors for each family $\mathbf{C}_a$. The region maps giving the desired difference vectors are formed as follows (for brevity sake, the positions i and j in the clique (i, j) are called the initial and the final one, respectively).

- (*i*) Let all the cliques $(i, j) \in \mathbf{C}_a$ in the map l contain only the labels k, that is, $l(i) = l(j) = k$, and let the map l′ differ by only one clique (i', j') such that $l(i) =$

k; $l(j) \neq k$ and the pixel j' is on the forward lattice border (in other words, there is no clique with the initial position in the pixel j'). Then, for $k = 0, \ldots, k_{\max}$ and these map pairs, the difference histogram subvectors are the same as the basis subvectors $\{\mathbf{b}_{a,k} : k = 0, \ldots, k_{\max}\}$.

* (*ii*) Let all the cliques $(i, j) \in \mathbf{C}_a$ in the map l contain only the labels k, that is, $l(i) = l(j) = k$, except for k separate pairs of the concatenated cliques (i_β, j_β); (j_β, m_β); $\beta = 0, \ldots, k-1$, such that the pixel i_β is on the backward lattice border (so that there are no cliques with the final positions in the pixels i_β) and $l(i_\beta) = l(j_\beta) = \beta$; $l(m_\beta) = k$. Also, let all the cliques in the map l' contain only the labels k except for k cliques $(i_\beta, j_\beta) : \beta = 0, \ldots, k-1$, such that the pixel j_β is on the forward lattice border (so that there are no cliques with the initial positions in the pixels j_β) and $l(i_\beta) = k$; $l(j_\beta) = \beta$. Then, for $k = 1, \ldots, k_{\max}$ and these map pairs, the difference histogram subvectors are the same as the basis subvectors $\{\mathbf{c}_{a,k} : k = 1, \ldots, k_{\max}\}$.

Therefore, all the basis subvectors from Table 2 for any second–order clique family take part in the non–zero difference vectors $\mathbf{\Delta}(\ldots)$ and, hence, the centered RLCH vectors are affinely independent in the vector subspace $\mathcal{S}$. This proof can also be adapted to the conditional models (17) and (18) of the region maps under a given image.

4.2 MLE of the Gibbs potentials

The above Lemmas show that the GPDs for the GRFs with multiple pairwise pixel interactions are strictly log-concave, that is, strongly unimodal, with respect to the potentials $\mathbf{V}$ (see BARNDORFF-NIELSEN [1978], JACOBSEN [1989]). In other words, the corresponding log-likelihood functions

$$L(\mathbf{V}|\mathbf{g}^{\circ}) = \frac{1}{|\mathbf{R}|} \ln\left(\Pr(\mathbf{g}^{\circ}|\mathbf{V})\right) \tag{20}$$

or

$$L(\mathbf{V}|\mathbf{g}^{\circ}, \mathrm{l}^{\circ}) = \frac{1}{|\mathbf{R}|} \ln\left(\Pr(\mathbf{g}^{\circ}, \mathrm{l}^{\circ}|\mathbf{V})\right) \tag{21}$$

of the potential vector $\mathbf{V}$ for a training sample $\mathbf{g}^{\circ}$ or training pair $(\mathbf{g}^{\circ}, \mathrm{l}^{\circ})$, respectively, are strictly concave. As shown by BARNDORFF-NIELSEN ([1978]), the likelihood function (20) has the unique finite maximum or, in other words, the MLE of the potentials exists if

and only if the following conditions hold for the marginal sample frequencies $F(q|\mathbf{g}^\circ) = \frac{H(q|\mathbf{g}^\circ)}{|\mathbf{R}|}$ and $F_a(d|\mathbf{g}^\circ) = \frac{H_a(d|\mathbf{g}^\circ)}{|\mathbf{C}_a|}$ of the gray levels and gray level differences, respectively:

$$\forall_{q\in\mathbf{Q}}\ \ 0 < F(q|\mathbf{g}^\circ) < 1; \quad \forall_{a\in\mathbf{A};\, d\in\mathbf{D}}\ \ 0 < F_a(d|\mathbf{g}^\circ) < 1. \tag{22}$$

The similar conditions for the marginal sample frequencies $F(q,k|\mathbf{g}^\circ,\mathrm{l}^\circ) = \frac{H(q,k|\mathbf{g}^\circ,\mathrm{l}^\circ)}{|\mathbf{R}|}$ and $F_{a,\alpha}(d,k|\mathbf{g}^\circ,\mathrm{l}^\circ) = \frac{H_{a,\alpha}(d,k|\mathbf{g}^\circ,\mathrm{l}^\circ)}{|\mathbf{C}_a|}$ regarding the finite maximum of (21), are as follows:

$$\forall_{q\in\mathbf{Q};\, k\in\mathbf{K}} \quad 0 < F(q,k|\mathbf{g}^\circ,\mathrm{l}^\circ) < 1; \tag{23}$$

$$\forall_{a\in\mathbf{A};\, d\in\mathbf{D};\, k\in\mathbf{K};\, \alpha\in\{0,1\}} \quad 0 < F_{a,\alpha}(d,k|\mathbf{g}^\circ,\mathrm{l}^\circ) < 1. \tag{24}$$

The conditions (22) and (24) are not too restrictive in practice because well-known Bayesian sample frequency estimates of the marginal probabilities never reach the limit bounds 0 or 1 and can be substituted for the above marginal frequencies.

Specific features of the likelihood functions (20) and (21), shown by GIMEL'FARB [1996a], [1996c], allow for learning both the interaction structure and Gibbs potentials from a given training sample or pair. This technique, briefly outlined below for the likelihood function (20), is easily adapted to the function (21), both for the models (13), (16) and for the models (17), (18).

Maximum Likelihood. The maximum of the log-likelihood function, given the centering (5), is in the point $\mathbf{V}^* \in \mathbf{S}$ where the gradient of this function is equal to zero. The components of the gradient are as follows (notice that below we use the centered frequencies and probabilities but just the same relations are valid with the initial non-centered ones, too):

$$\begin{aligned}
\frac{\partial L(\mathbf{V}|\mathbf{g}^\circ)}{\partial V(q)} &= F_{\mathrm{cn}}(q|\mathbf{g}^\circ) - \mathcal{E}\{F_{\mathrm{cn}}(q|\mathbf{g})|\mathbf{V}\} \equiv F_{\mathrm{cn}}(q|\mathbf{g}^\circ) - M_{\mathrm{cn}}(q|\mathbf{V});\\
\frac{\partial L(\mathbf{V}|\mathbf{g}^\circ)}{\partial V_a(d)} &= \rho_a \cdot (F_{\mathrm{cn},a}(d|\mathbf{g}^\circ) - \mathcal{E}\{F_{\mathrm{cn},a}(d|\mathbf{g})|\mathbf{V}\})\\
&\equiv \rho_a \cdot (F_{\mathrm{cn},a}(d|\mathbf{g}^\circ) - M_{\mathrm{cn},a}(d|\mathbf{V}))
\end{aligned} \tag{25}$$

so that the gradient lies in the subspace $\mathcal{S}$ of the centered potential and histogram vectors, too. Here, $M_{\ldots}(\ldots|\mathbf{V}) \equiv \mathcal{E}\{\ldots|\mathbf{V}\}$ is the centered marginal probability, or the expectation of a centered marginal sample frequency, under the GPD of (4) with the potential vector $\mathbf{V}$, and the factor $\rho_a = |\mathbf{C}_a|/|\mathbf{R}|$.

The following system of equations:

$$\forall_{q\in\mathbf{Q}}\ F_{\mathrm{cn}}(q|\mathbf{g}^{\circ}) = M_{\mathrm{cn}}(q|\mathbf{V}^{*});\ \ \forall_{d\in\mathbf{D};\,a\in\mathbf{A}}\ F_{\mathrm{cn},a}(d|\mathbf{x}^{\circ}) = M_{\mathrm{cn},a}(d|\mathbf{V}^{*}) \tag{26}$$

holds at the unique maximum point of the likelihood function. The samples $\mathbf{g}$ having a given GPD can be generated by well-known pixel-wise stochastic relaxation techniques (METROPOLIS et al. [1953], CROSS and JAIN [1983], GEMAN and GEMAN [1984], CHELLAPPA and JAIN [1993]). This makes possible (YOUNES [1988]) to find the desired MLE by solving the system (26) with a stochastic approximation starting from a first approximation of the potentials.

The Analytic First Approximation of the MLE $\mathbf{V}^{*}$ is derived in (GIMEL'FARB [1996a]) by expanding the log-likelihood function into a truncated Taylor's series about the zero point $\mathbf{V} = \mathbf{0}$. Under the involved potential centering, this point corresponds to the IRF. The expansion is maximized then along the gradient in this point. For the GPD (3) these estimates are as follows:

$$\begin{aligned} \forall_{q\in\mathbf{Q}} \qquad & V_{[0]}(q) = \lambda_{[0]}\cdot D_{[0]}(q); \\ \forall_{d\in\mathbf{D};\,a\in\mathbf{A}} \quad & V_{a,[0]}(d) = \lambda_{[0]}\cdot\rho_a\cdot D_{a,[0]}(d) \end{aligned} \tag{27}$$

where $D_{[0]}(q) = F_{\mathrm{cn}}(q|\mathbf{g}^{\circ}) - M_{\mathrm{cn,irf}}(q)$ and $D_{a,[0]}(d) = F_{\mathrm{cn},a}(d|\mathbf{g}^{\circ}) - M_{\mathrm{cn,dif}}(d)$. The non-centered and centered marginal probabilities of the signals and signal differences for the IRF have the following forms:

$$\begin{aligned} \forall_{q\in\mathbf{Q}} \qquad & M_{\mathrm{irf}}(q) = \frac{1}{1+q_{max}}\ ; \qquad M_{\mathrm{cn,irf}}(q) = 0; \\ \forall_{d\in\mathbf{D};\,a\in\mathbf{A}} \quad & M_{\mathrm{dif}}(d) = \frac{1+q_{\max}-|d|}{(1+q_{\max})^2}\ ; \quad M_{\mathrm{cn,dif}}(d) = M_{\mathrm{dif}}(d) - \frac{1}{1+2q_{max}} \end{aligned} \tag{28}$$

The scaling factor $\lambda_{[0]}$ in (27) is computed from the known marginals as follows:

$$\lambda_{[0]} = \frac{\displaystyle\sum_{q\in\mathbf{Q}} D^2_{[0]}(q) + \sum_{a\in\mathbf{A}}\rho_a^2\sum_{d\in\mathbf{D}} D^2_{a,[0]}(d)}{\displaystyle\sum_{q\in\mathbf{Q}}\sigma_{\mathrm{irf}}(q)\cdot D^2_{[0]}(q) + \sum_{a\in\mathbf{A}}\rho_a^3\sum_{d\in\mathbf{D}}\sigma_{\mathrm{dif}}(d)\cdot D^2_{a,[0]}(d)} \tag{29}$$

Here, $\sigma_{\mathrm{irf}}(q) = M_{\mathrm{irf}}(q)\cdot(1 - M_{\mathrm{irf}}(q))$ and $\sigma_{\mathrm{dif}}(d) = M_{\mathrm{dif}}(d)\cdot(1 - M_{\mathrm{dif}}(d))$ are the variances of the marginal frequencies. It is worth noting that the larger the lattice, the closer the factors ρ_a to unity. Therefore, the estimates in (27) and (29) are almost independent of the lattice size. The GPD in (3) with $|\mathbf{A}|$ clique families involves $G = q_{\max}\cdot(2\cdot|\mathbf{A}| + 1)$ potential values to be estimated from $|\mathbf{R}|$ signals for a training sample. To ensure the asymptotic consistency of the MLEs, it is necessary that $G \ll |\mathbf{R}|$.

Search for the Interaction Structure. The analytic initial estimates of the potentials show that the relative interaction strength for each clique family $\mathbf{C}_a$ in the models (1) and (3) can be represented, for instance, by a chi-square distance between the marginal GLD frequencies for the training sample and for the IRF (GIMEL'FARB [1996a]). The smaller the distance (that is, the closer the potential estimates to the zero point), the weaker the interaction. Therefore, the clique families with a sufficiently weak interaction strength can be excluded from the models or, what is the same, the potential values for them can be set to zero.

Instead of the chi-square distances, the *relative Gibbs energies* for the training sample, showing contributions of each clique family to the total exponent of the GPD (1) or (3), can directly be used to compare the interaction strengths for the clique families. Let

$$E_{a,[0]}(\mathbf{g}^\circ) = \rho_a \sum_{d\in\mathbf{D}} \left(F_{\mathrm{cn},a}(d|\mathbf{g}^\circ) - M_{\mathrm{cn,dif}}(d)\right)\cdot F_{\mathrm{cn},a}(d|\mathbf{g}^\circ) \tag{30}$$

denote a relative Gibbs energy of the clique family a in the sample $\mathbf{g}^\circ$, given the initial potential estimates (27). Let $\mathbf{W}$ be a given search set containing all the clique families within a given large range of possible intra-clique shifts: $\mathbf{W} = \{(\mu,\nu) : |\mu| \le \mu_{\max};\ |\nu| \le \nu_{\max}\}$. A rich variety of the clique families, defined by the search set $\mathbf{W}$, are compared by their relative Gibbs energies $E_{a=(\mu,\nu),[0]}(\mathbf{g}^\circ)$ computed for a given training sample $\mathbf{g}^\circ$. In so doing, the energy values over the search set $\mathbf{W}$ are represented as a 2D energy function $E(\mu,\nu|\mathbf{g}^\circ) = \left\{\rho_a E_{a,[0]}(\mathbf{g}^\circ) :\ a=(\mu,\nu)\in\mathbf{W}\right\}$ with planar Cartesian coordinates (μ,ν). This representation constitutes an *interaction map* which shows relative contributions of each of the clique family to the total energy and can be displayed, for a visual analysis, in a grayscale or color form (GIMEL'FARB [1996a], GIMEL'FARB and JAIN [1996]).

A following simple technique, first proposed by GIMEL'FARB [1996a], allows to find most characteristic clique families comprising the desired interaction structure of the models (1) and (3). It is based on a direct thresholding of the interaction map:

$$\mathbf{A} = \left\{a :\ a\in\mathbf{W};\ E_{a,[0]}(\mathbf{g}^\circ) > \theta\right\} \tag{31}$$

where θ denotes a given threshold. In (GIMEL'FARB [1996a] – [1996c]) it was chosen as a function either of the mean relative energy $\overline{E}$ and standard deviation σ_E in the interaction map: $\theta = \overline{E} + c\cdot\sigma_E$, where $c = 3\ldots 4$, or of the maximum relative energy: $\theta = c\cdot E_{\max}$ where $c = 0.25\ldots 0.35$. But, further theoretical investigation is needed for optimizing such a search.

Stochastic Approximation Refinement of the Potentials. After finding most characteristic interaction structure, the initial estimates (27) for the chosen clique families (31) are refined by the stochastic approximation techniques (YOUNES [1988]). They are based on generating, by stochastic relaxation, a Markov chain of the model samples under a gradually changing GPD. Let a *macrostep* mean one pass of the stochastic relaxation round the total lattice $\mathbf{R}$ without repetition of the pixels. At each (macro)step t, the current image $\mathbf{g}^{[t]}$ is generated from a previous one $\mathbf{g}^{[t-1]}$ under the current GPD $\Pr(\mathbf{g}|\mathbf{V}_{[t-1]})$. Potential estimates $\mathbf{V}_{[t]}$ are updated using the GLH and GLDHs for the current generated image in line with the differences between the marginal frequencies for the training sample and generated samples:

$$\begin{aligned} \forall_{q\in\mathbf{Q}} \qquad & V_{[t]}(q) = V_{[t-1]}(q) + \lambda_{[t]} \cdot \Big(F(q|\mathbf{g}^{\circ}) - F(q|\mathbf{g}_{[t]})\Big) \\ \forall_{d\in\mathbf{D};\ a\in\mathbf{A}} \quad & V_{a,[t]}(d) = V_{a,[t-1]}(d) + \lambda_{[t]} \cdot \rho_a \cdot \Big(F_a(d|\mathbf{g}^{\circ}) - F_a(d|\mathbf{g}_{[t]})\Big) \end{aligned} \tag{32}$$

Here, t is the number of the approximation macrostep ($t = 0$ for the initial estimates (27)), $\mathbf{g}_{[t]}$ is the sample generated under the GPD $\Pr\Big(\mathbf{g}|\mathbf{V}_{[t-1]}\Big)$ by stochastic relaxation (notice that $\mathbf{g}_{[0]}$ is a sample of the IRF), and the scaling factor $\lambda_{[t]}$ determines a contracted step along the current approximation of the gradient (26).

The scaling factor $\lambda_{[t]}$ decreases from the starting value $\lambda_{[0]}$ in (29) as $\dfrac{c_0+1}{c_1+c_2\cdot t}$. Theoretically justified choice of the control parameters $c_0,\ c_1,\ c_2$ giving almost sure convergence of the updating process (32) to the desired MLE of the potentials is given by YOUNES [1988] who also has shown that such a choice yields too slow convergence to the desired estimates and should be replaced by some empirically found values. Such a heuristic choice of YOUNES [1988], slightly modified in (GIMEL'FARB [1996a]), is used in the experiments presented in (GIMEL'FARB [1996a] – [1996c]) and in Section 7 below.

This stochastic approximation refinement of the potentials MLE needs a sizable number of the (macro)steps for ensuring the convergence to the desired maximum of the likelihood function, that is, to the solution of the system (26). Thus, in practice, it is sometimes difficult to implement the above learning scheme.

It is easily seen that the joint and conditional models of the grayscale images and region maps in Section 3 have learning procedures that are quite similar to the above-mentioned one. The learnt interaction structure and initial analytic estimates of the Gibbs potentials allow to simulate homogeneous and piecewise-homogeneous image textures and to segment the latter ones by using the approximation to a Bayesian decision framework proposed in

(GIMEL'FARB [1996c]).

It is worthy to note that the above GPDs are closely similar to the δ–function in that the images with the significantly non-zero probabilities form a very small subset concentrated around the maximum probable image(s). Therefore, the stochastic approximation processes of (32) can also be regarded as an adaptive image generating technique called in (GIMEL'FARB [1996a, 1996c]) a *Controllable Simulated Annealing* (CSA).

4.3 Controllable simulated annealing

The CSA differs from the stochastic approximation refinement of the initial potential values only in its final aim to get the images themselves instead of the model parameters. At each CSA (macro)step, the samples are generated by a pixelwise stochastic relaxation technique, say, by an algorithm of METROPOLIS et al. [1953] or by a Gibbs sampler (GEMAN and GEMAN [1984]) and the potentials are changing just as in the stochasic approximation solution of the equations (26). But, here we exploit a by-product of this solution, namely, a close proximity between the GLDHs of the training sample and the GLDHs of the samples generated with the refined potentials. Under the GPD $\Pr(\mathbf{g}|\mathbf{V}^*)$ with the MLE $\mathbf{V}^*$ of the potentials, the expected total Gibbs energy over the parent population is equal to the total energy for the training sample (see the equations (26) relating the training and expected marginals for this MLE):

$$\mathcal{E}\{\mathbf{V}^* \bullet \mathbf{H}(\mathbf{g})\} = \mathbf{V}^* \bullet \mathcal{E}\{\mathbf{H}(\mathbf{g})\} = \mathbf{V}^* \bullet \mathbf{H}(\mathbf{g}^\circ)$$

and, usually, the variance of the total energy is rather low. This ensures the high probabilities of the simulated images in relation to the used GPDs. In the case of image simulating, such a proximity yields, by and large, a fairly good visual similarity between the generated and goal homogeneous image textures. The finally obtained samples approach more closely the training ones, as regarding the signal histograms, than the samples that are generated by using the usual stochastic relaxation with the fixed learnt potentials. Also, the CSA gives much more freedom in choosing the control parameters of the stochastic approximation because one needs only the proximity between the training and the final generated sample marginals and no true convergence to the potential MLE itself.

The stochastic relaxation technique for the non-Markov Gibbs model (3) differ somewhat from the one for the Markov/Gibbs model (1). In the Markov/Gibbs model case, each step of the pixelwise relaxation involves the summation of the potentials only over the local neighborhood of the current pixel. This neighborhood contains all the neighbors of the

pixel (that is, the local neighborhood is the "star-like" union of all the cliques containing the current pixel). The non-Markov models preserve the like computations for all the pixels, except for the pixels with a solitary maximum or minimum gray level. Only in the latter rare case, the actual neighborhood of the pixel is the whole lattice and potentials are summed up over the lattice. As a result, the computational complexity of the relaxation does not increase substantially.

The CSA differs from the usual simulated annealing (GEMAN and GEMAN [1984], CHELLAPPA and JAIN [1993]) in that the system (26) presents an explicit unimodal measure of the proximity between the current generated sample and the goal training one in terms of their histograms. Also, the process of approaching the maximum proximity is equivalent to the one of maximizing the unimodal likelihood function and this results in a good convergence of the CSA. When using the CSA, we avoid the fairly long potential refinement stage and can simulate the model samples just after choosing the characteristic interaction structure.

4.4 Bayesian framework to simulate and segment textures

The above models embed both simulation of the homogeneous and piecewise-homogeneous textures and segmentation of the piecewise-homogeneous textures into a unified Bayesian processing framework. It is based on a stochastic relaxation generation of the image samples under a given Gibbs model (GIMEL'FARB [1996c]).

Let $\mathbf{x}$ denote a data sample (that is, an image, a region map, or an image–map pair). Let $\mathbf{X}$ be the parent population of the samples. Let $\mathbf{Y}$ represent fixed parameters of the GPD. Also, let n denote a signal (a gray level, a region label, or their pair) in the pixel and $\mathbf{N}$ be a finite set of the signal values. The Bayesian framework implements either the simple MAP-decision with the maximum a posteriori probability of the desired sample:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathbf{X}} \Pr(\mathbf{x}|\mathbf{Y}), \tag{33}$$

or the compound MPM-decision with maximum posterior marginal probabilities of the sample components (this latter one was first introduced by ABEND [1966]):

$$\forall_{i\in\mathbf{R}} \;\; \mathbf{x}^*(i) = \arg\max_{\mathbf{n}\in\mathbf{N}} \Pr_i(\mathbf{x}(i) = n|\mathbf{Y}). \tag{34}$$

The MAP-decision involves the well-known *simulated annealing* techniques searching for most probable sample by changing the Gibbs potentials as to stay in a vicinity of the

desired maximum during shrinking the GPD from the IRF with equiprobable samples up to the δ-function in the limit, when the number of the (macro)steps approaches the infinity (see GEMAN and GEMAN [1984], CHELLAPPA and JAIN [1993], YOUNES [1988] for more details). The MPM-decision exploits the stochastic relaxation with the fixed potentials to get a set of the samples under the given GPD for estimating the marginals. But, the generated Markov chain of the images has to reach an equilibrium state and one can ensure this only in the limit.

Thus, both the cases need, in principle, many relaxation steps. Theoretically justified simulated annealing schedules of changing the potentials are derived as to ensure the convergence to the maximum point even in the worst cases rarely met in practice. Thus, usually they are replaced by some empirically found schedules but then it is hard to say whether the final decision is really the MAP-one. The MPM-case needs relatively small number of steps to get good frequency estimates of the marginals if the Markov chain of the generated images is really in equilibrium. But, in practice, it is hard to verify whether the chain has reached this state.

In most cases, the introduced GPDs are close to the δ-function with respect to the total Gibbs energies of the images. Therefore, the following conjecture seems to be valid:

Conjecture 1 *The images giving the MAP- or the MPM-decision have the partial Gibbs energies and, therefore, possess particular signal histograms which differ little from the corresponding energies and histograms for the training sample(s).*

Under this conjecture, the CSA allows to approximate the desired Bayesian decisions by generating the images with the desired signal histograms, say, with the GLH, or RLH, or joint GL/RLH and with a selected subset of the GLDHs, or RLCHs, or joint GLD/RLCHs, which are close to the given training histograms.

5 Conditional MLE of the Gibbs Potentials

The foregoing learning scheme involves a sizable number (G) of the unknown potential values to be computed for the models (1) and (3) by stochastic approximation. This number can be reduced to only $|\mathbf{A}| + 1$ unknown parameters by exploiting, instead of the unconditional MLE (22), the conditional one provided that the training sample $\mathbf{g}$ may rank the top place attainable within the parent population $\mathbf{G}$ in the Gibbs energy (30).

5.1 Feasible top rank principle

Let $e(\mathbf{g}|\mathbf{v}) = \sum_{q\in\mathbf{Q}} V(q)\cdot H_{\mathrm{cn}}(q|\mathbf{g})$ be the partial energy for the first-order, that is, pixel-wise clique family and $e_a(\mathbf{g}|\mathbf{v}_a) = \sum_{d\in\mathbf{D}} V_a(d)\cdot H_{\mathrm{cn},a}(d|\mathbf{g})$ denote the partial energy for the second-order, or pairwise clique family $\mathbf{C}_a$. Here, $\mathbf{v} = \{V(q) : \ q \in \mathbf{Q}\}$ and $\mathbf{v}_a = \{V_a(d) : \ d \in \mathbf{D}\}$ denote the corresponding potential subvectors. The total Gibbs energy (for definiteness, in the model (4)) is as follows: $E(\mathbf{g}|\mathbf{V}) \equiv \mathbf{V}\bullet\mathbf{H}_{\mathrm{cn}}(\mathbf{g}) = e(\mathbf{g}|\mathbf{v}) + \sum_{a\in\mathbf{A}} e_a(\mathbf{g}|\mathbf{v}_a)$.

The following Lemma holds for the Gibbs potentials:

Lemma 3 *Let the samples $\mathbf{g}\in\mathbf{G}$ are ranked in ascending order of the total Gibbs energies $E(\mathbf{g}|\mathbf{V})$ for each the potential vector $\mathbf{V}$. The vector $\mathbf{V}^{\circ}(\mathbf{\Lambda})$ that ranks the training sample $\mathbf{g}^{\circ}$ to a feasible top place within the parent population $\mathbf{G}$ in the total energy possesses the following explicit, except for scaling factors, form:*

$$\mathbf{V}^{\circ}(\mathbf{\Lambda}) = \{\lambda\cdot F_{\mathrm{cn}}(q|\mathbf{g}^{\circ}) : \ q\in\mathbf{Q};\ \lambda_a\cdot F_{\mathrm{cn},a}(d|\mathbf{g}^{\circ}) : \ d\in\mathbf{D};\ a\in\mathbf{A}\}\,. \tag{35}$$

Here, $\mathbf{\Lambda} = \{\lambda,\lambda_a : \ a\in\mathbf{A};\ \lambda,\lambda_a>0\}$ denotes a vector of the arbitrary positive scaling factors.

To prove the Lemma 3, let us notice that, for every clique family, the sample ranking in the partial energy is invariant to the potential (and energy) normalization that reduces the corresponding potential subvector $\mathbf{v}_{...}$ to the unit subvector $\tilde{\mathbf{v}}_{...} = \mathbf{v}_{...}/|\mathbf{v}_{...}|$. Let $\mathbf{F}_{\mathrm{cn}}(\mathbf{g}^{\circ}) = \{F_{\mathrm{cn}}(q|\mathbf{g}^{\circ}) : \ q\in\mathbf{Q}\}$ and $\mathbf{F}_{\mathrm{cn},a}(\mathbf{g}^{\circ}) = \{F_{\mathrm{cn},a}(d|\mathbf{g}^{\circ}) : \ d\in\mathbf{D}\}$ denote, respectively, the centered vectors of the marginal gray level frequencies and of the marginal gray level difference frequencies for the clique family $\mathbf{C}_a$. It is easy to see that the unit subvectors $\tilde{\mathbf{v}}^{\circ} = \dfrac{\mathbf{F}_{\mathrm{cn}}(\mathbf{g}^{\circ})}{|\mathbf{F}_{\mathrm{cn}}(\mathbf{g}^{\circ})|}$ and $\tilde{\mathbf{v}}_a^{\circ} = \dfrac{\mathbf{F}_{\mathrm{cn},a}(\mathbf{g}^{\circ})}{|\mathbf{F}_{\mathrm{cn},a}(\mathbf{g}^{\circ})|}$ maximize the normalized partial energy $e(\mathbf{g}^{\circ}|\tilde{\mathbf{v}})$ and the normalized partial energy $e_a(\mathbf{g}^{\circ}|\tilde{\mathbf{v}}_a)$, respectively.

Every arbitrary potential subvector obtained by scaling such a unit subvector, ranks the training sample $\mathbf{g}^{\circ}$ in the corresponding partial energy to the same top place which may be feasible among the samples $\mathbf{g}\in\mathbf{G}$ of the parent population as compared to any other potential subvector. Therefore, the potential estimates (35) give the feasible top rank for the training sample in the total energy summed over all the families.

In other words, in spite of the changes of the partial energies for the different potentials $\mathbf{V}$, the training sample may occupy the feasible top rank in the total Gibbs energy if the potentials are proportional to the centered sample marginals. Therefore, the following Lemma is valid:

Lemma 4 *The conditional MLE of the potentials*

$$\begin{aligned}\mathbf{V}^{\circ}(\mathbf{\Lambda}^{\star}) \;\; = \;\; & \{V^{\star}(q)=\lambda^{\star}\cdot F_{\mathrm{cn}}(q|\mathbf{g}^{\circ}):\; q\in\mathbf{Q};\\ & V^{\star}_{a}(d)=\lambda^{\star}_{a}\cdot F_{\mathrm{cn},a}(d|\mathbf{g}^{\circ}):\; d\in\mathbf{D};\; a\in\mathbf{A}\}\end{aligned} \tag{36}$$

such that $\mathbf{\Lambda}^{\star}=\arg\max_{\mathbf{\Lambda}}\Pr(\mathbf{g}^{\circ}|\mathbf{V}^{\circ}(\mathbf{\Lambda}))$ *yields the maximum probability of the training sample* $\mathbf{g}^{\circ}$ *provided that it may occupy the top rank which is feasible for this sample within the parent population ordered in the total Gibbs energies.*

Therefore, the introduced *feasible top rank principle* produces the explicit conditional potential MLEs (36) with only the scaling factors $\Lambda^{\star}$ to be computed for each clique family by maximizing the likelihood function.

5.2 Estimation of the scaling factors

The desired factors are learnt in a similar way as the potentials themselves in Section 4: first, by approximating them analytically and finding the characteristic interaction structure, and then, by refining them for the chosen clique families with the stochastic approximation.

Generally, the estimate (36) may differ from the true unconditional MLE (26). But, for the GPDs in (1) - (3) some plausible considerations exist that both the estimates, at least, are fairly close if not equivalent. This conjecture needs further theoretical investigations. The supporting considerations are based on a close similarity between the analytic first approximation (27) of the unconditional potential MLE and the conditional MLE (36) and on the fixed ranking of the samples under the potential scaling (35) and, by symmetry, under a uniform scaling of the sample histograms.

The Analytic First Approximation of the factors is obtained by the same technique of a truncated Taylor's series expansion of the likelihood function about the zero point $\mathbf{\Lambda}=\mathbf{0}$ as in Section 4. The approximation is as follows:

$$\lambda_{[0]}=\alpha_{[0]}\cdot E_{[0]}(\mathbf{g}^{\circ});\;\forall_{a\in\mathbf{A}}\;\lambda_{a,[0]}=\alpha_{[0]}\cdot E_{a,[0]}(\mathbf{g}^{\circ}) \tag{37}$$

where $E_{[0]}(\mathbf{g}^\circ) = \sum_{q\in\mathbf{Q}} F^2_{\mathrm{cn}}(q|\mathbf{g}^\circ)$ is the relative first-order Gibbs energy and $E_{a,[0]}(\mathbf{g}^\circ)$ is the second-order one of (30). The factor $\alpha_{[0]}$ is computed from these energies as:

$$\alpha_{[0]} = \frac{E^2_{[0]}(\mathbf{g}^\circ) + \sum_{a\in\mathbf{A}} E^2_{a,[0]}(\mathbf{g}^\circ)}{E^2_{[0]}(\mathbf{g}^\circ)\cdot U_{[0]}(\mathbf{g}^\circ) + \sum_{a\in\mathbf{A}} E^2_{a,[0]}(\mathbf{g}^\circ)\cdot U_{a,[0]}(\mathbf{g}^\circ)} \ . \tag{38}$$

Here,

$$\begin{aligned} U_{[0]}(\mathbf{g}^\circ) &= \sum_{q\in\mathbf{Q}} \Big(F^2_{\mathrm{cn}}(q|\mathbf{g}^\circ)\cdot\sigma_{\mathrm{irf}}(q)\Big) \\ U_{a,[0]}(\mathbf{g}^\circ) &= \rho_a \sum_{d\in\mathbf{D}} \Big(F^2_{\mathrm{cn},a}(d|\mathbf{g}^\circ)\cdot\sigma_{\mathrm{dif}}(d)\Big) \end{aligned} \tag{39}$$

Here, as in (29), $\sigma_{\mathrm{irf}}(q)$ and $\sigma_{\mathrm{dif}}(d)$ denote the variances of the marginal frequencies of the gray level and gray level differences for the IRF.

Search for the Interaction Structure exploits the interaction map which is formed in this case by using the weighted energies (30):

$$\mathbf{E}_{[0]}(\mathbf{g}^\circ) = \Big\{\rho_a\cdot\omega_{a,[0]}\cdot E_{a,[0]}(\mathbf{g}^\circ):\ a\in\mathbf{A_W}\Big\}$$

where the weight $\omega_{a,[0]} = \sum_{d\in\mathbf{D}} F^2_{\mathrm{cn},a}(d|\mathbf{g}^\circ)$.

Stochastic Approximation Refinement of the Factors also exploits the similar energies which depend on the proximity between the marginal gray level difference frequencies for each clique family in the training and generated samples. At each step t of the stochastic approximation, the current factors are updated as follows:

$$\begin{aligned} \lambda_{[t+1]} &= \lambda_{[t]} + \alpha_{[t]}\cdot E_{[t]}(\mathbf{g}^\circ,\mathbf{g}_{[t]}); \\ \forall_{a\in\mathbf{A}}\quad \lambda_{a,[t+1]} &= \lambda_{a,[t]} + \alpha_{[t]}\cdot E_{a,[t]}(\mathbf{g}^\circ,\mathbf{g}_{[t]}) \end{aligned} \tag{40}$$

where $\mathbf{g}_{[t]}$ is the sample generated at this step, $\alpha_{[t]}$ is the current scaling factor, and

$$\begin{aligned} E_{[t]}(\mathbf{g}^\circ,\mathbf{g}_{[t]}) &= \sum_{q\in\mathbf{Q}} \Big(F_{\mathrm{cn}}(q|\mathbf{g}^\circ) - F_{\mathrm{cn}}(q|\mathbf{g}_{[t]})\Big)\cdot F_{\mathrm{cn}}(q|\mathbf{g}^\circ); \\ E_{a,[t]}(\mathbf{g}^\circ,\mathbf{g}_{[t]}) &= \rho_a \sum_{d\in\mathbf{D}} \Big(F_{\mathrm{cn},a}(d|\mathbf{g}^\circ) - F_{\mathrm{cn},a}(d|\mathbf{g}_{[t]})\Big)\cdot F_{\mathrm{cn},a}(d|\mathbf{g}^\circ). \end{aligned} \tag{41}$$

6 Quantitative Features of Self-Similarity

Approximation of the natural texture by the introduced Gibbs models allows to specify more in detail the self-similarity concept and give quantitative definitions to some basic texture features in terms of the GLH and GLDHs as the sufficient statistics of the model.

6.1 Evaluating the size of the texture patch

The marginal probabilities of the gray levels and gray level differences are assumed to be spatially invariant in a given homogeneous texture. Thus, the minimum size of a texture patch can be deduced from the desired absolute or relative precision of approximating these probabilities by the collected histograms, in particular, by the GLH and the GLDHs for the normalized image $\mathbf{g}^{\circ,\mathrm{rf}}$ in the case of the non-Markov Gibbs model (3). Actually, the gray level differences in the different cliques of the same family are statistically dependent but these dependencies are not clearly understood. One can exploit the Chebyshev's inequality to estimate, with a large margin, the histogram size $|\mathbf{C}_a|$ which yields the maximum absolute error ε of the probability estimates with a confidence level $1-\alpha$: $|\mathbf{C}_a| \geq \frac{0.25}{\varepsilon^2 \cdot \alpha}$. In other words, under this histogram size, the absolute difference between the actual marginal probability and its estimate from the histogram is greater than a given threshold ε only with a small probability which is not greater than α.

If we consider the square texture patch $n \times n$ then its rough minimum linear size is as follows: $n = \max\{\mu_{\max},\, \nu_{\max}\} + \frac{0.5}{\varepsilon \cdot \sqrt{\alpha}}$. Here, $\mu_{\max}$ and $\nu_{\max}$ specify the maximum horizontal and vertical shift between the pixels in the cliques, respectively. Let, for example, $\alpha = 0.01$ (that is, the confidence level is 0.99), $\varepsilon = 0.05$, and $\mu_{\max} = \nu_{\max} = 40$. In this case, the patch linear size is as follows: $n = 140$.

6.2 Measuring the self-similarity

Now, the spatial self-similarity in the texture can be defined as the similarity between the GLDHs for the different patches to within a given set U of the admissible scales and orientations or, more generally, of the projective transformations of these patches. The total chi-square distance between the GLDHs for the corresponding clique families in the patches to be compared can be used as a quantitative similarity measure:

$$D(\mathbf{g}_1^{\mathrm{rf}}, \mathbf{g}_2^{\mathrm{rf}}) = \sum_{a \in \mathbf{A}} \chi^2 \left(\mathbf{H}_a(\mathbf{g}_1^{\mathrm{rf}}), \mathbf{H}_a(\mathbf{g}_2^{\mathrm{rf}}) \right). \tag{42}$$

The correspondence of the clique families means here the same inter-pixel shifts μ_a, ν_a. To take account of the admissible scales and orientations of the patches, the distance (42)

has to be minimized over the given set $\mathbf{U}$ of such relative transformations. Experiments presented in (GIMEL'FARB and JAIN [1996]) show that the scale and orientation changes of the patch patterns result in the like changes of the interaction maps, that is, of the inter-pixel shifts defining the clique families. This allows for describing any transformation $u \equiv (\lambda, \phi) \in \mathbf{U}$ by a simple rearrangement of the initial GLDHs $\mathbf{H}_a(\mathbf{g})$; $a \in \mathbf{A}$, in line with the transformed shifts $(\mu_{a(u)}, \nu_{a(u)})$. Here,

$$\begin{aligned} \mu_{a(u)} &= \mathrm{Int}\left\{0.5 + \lambda \cdot (\mu_a \cdot \cos\phi + \nu_a \cdot \sin\phi)\right\}; \\ \nu_{a(u)} &= \mathrm{Int}\left\{0.5 + \lambda \cdot (-\mu_a \cdot \sin\phi + \nu_a \cdot \cos\phi)\right\}, \end{aligned} \tag{43}$$

where $\mathrm{Int}\{\ldots\}$ denotes the integer part of the real number.

One can simplify this self-similarity test by reducing the distance computation to the chosen interaction structure. In line with the used approximation, any GLDH which is absent in the models to be compared is replaced by the GLDH for the IRF. This allows to reject most dissimilar patches using only their structural similarity (that is, relative number of corresponding clique families with respect to their total number in both the patches) and only then exploit the above similarity measure in (42) to rank structurally similar patches (GIMEL'FARB and Jain [1996]).

7 Experimental Results and Conclusions

7.1 Supervised texture simulation

7.1.1 Interaction maps for natural and simulated textures

Figures 2 and 3 display the natural texture samples from BRODATZ [1966], simulated textures, and their interaction maps. Figure 2 shows, respectively, six digitized fragments 128×128 of the weakly homogeneous natural texture D3 (Reptile Skin) and the corresponding interaction maps representing 3280 clique families in the search window $\mathbf{W}$ formed with the parameters $\mu_{\max} = 40$; $\nu_{\max} = 40$. The shown map contains two square boxes (2×2 pixels in Figures 2 and 3) with relative coordinates $(\mu_a,\ \nu_a)$ and $(-\mu_a,\ -\nu_a)$ per family $\mathbf{C}_a$ with the intra-clique shift $(\mu_a,\ \nu_a) \in \mathbf{W}$. The origin of Cartesian coordinates (μ, ν) is marked by a white square. For visual representation, the energies $E_{a,[0]}(\mathbf{g}^{\circ})$ (see (30)) are coded by the gray levels: the darker these two boxes, the greater the energy and the stronger the interaction. Figure 3 shows the like fragments and interaction maps for the natural texture D14 (Woven Aluminum Wire). It is almost obvious that the presented textures possess the self-similarity in the introduced "model-based" sense if the texels are

sufficiently large to represent the repetitive parts of these patterns but, from a viewpoint of the translation invariant pairwise interactions, the self-similarity is rather weak. Nonetheless, we show below that the Gibbs model (3) reflects most characteristic features which are visually perceived in these textures.

The shown interaction maps exhibit relative contributions of different clique families presented in the search window to the imaginary Gibbs model that contains all these families. This allows for approximating a given texture type by a reduced model with zero-valued potentials for the families with too weak interaction strengths. Generally, this suggests that all the models have the same interaction structure that correspond to the largest possible search set $\mathbf{W}$ and differ only by the potentials: the non-zeroth values for the characteristic clique families and zeroth values for all the other families. It is this feature that simplifies comparisons of different textures, say, for the image retrieval (JAIN and GIMEL'FARB [1995], GIMEL'FARB and JAIN [1996]) and allows for expanding the model onto more complex piecewise-homogeneous textures outlined in Sections 3.1 and 3.2 (see also GIMEL'FARB [1996c]).

Here, a heuristic search for most characteristic clique families is done by thresholding the interaction map with the threshold $\theta = \overline{E} + c \cdot \sigma_E$ (see Section 4). The interaction structures in Figures 2 and 3, learnt with $c = 3$, have similar geometric forms for different samples which reflect basic visually perceived hexagonal or tetragonal patterns of these textures. The structures of most homogeneous samples approximately include the same numbers of clique families: say, $56 \ldots 60$ in Figure 2 (the samples T-D3a, T-D3c, T-D3d, T-D3f) and $71 \ldots 79$ in Figure 3 (the samples T-D14a – T-D14f). Local inhomogeneities affect the learnt structures. For instance, due to notable cell size changes in the sample T-D3b or to a non-homogeneous cell arrangement in the sample T-D3e (Figure 2), some characteristic long-range interactions are missing from the structures ST3b and ST3e.

7.1.2 Comparing natural and simulated textures

Figures 2 and 3 also demonstrate the homogeneous textured samples simulated by the CSA under the model (3) with the learnt interaction structures and potentials. The simulation starts from an IRF sample. The chi-square distances between the normalized GLH and GLDHs of the training and simulated samples are reduced from the starting values of $601,000 \ldots 1,329,000$ at the first macrostep ($t = 0$) to $1425 \ldots 6125$ at the last one ($t = 200$) for the samples in Figure 2 and from $965,000 \ldots 2,712,000$ to $2300 \ldots 7500$ for the samples in Figure 3. As a result, the interaction maps and learnt structures computed for the simulated samples closely match the initial ones for the training samples both in the

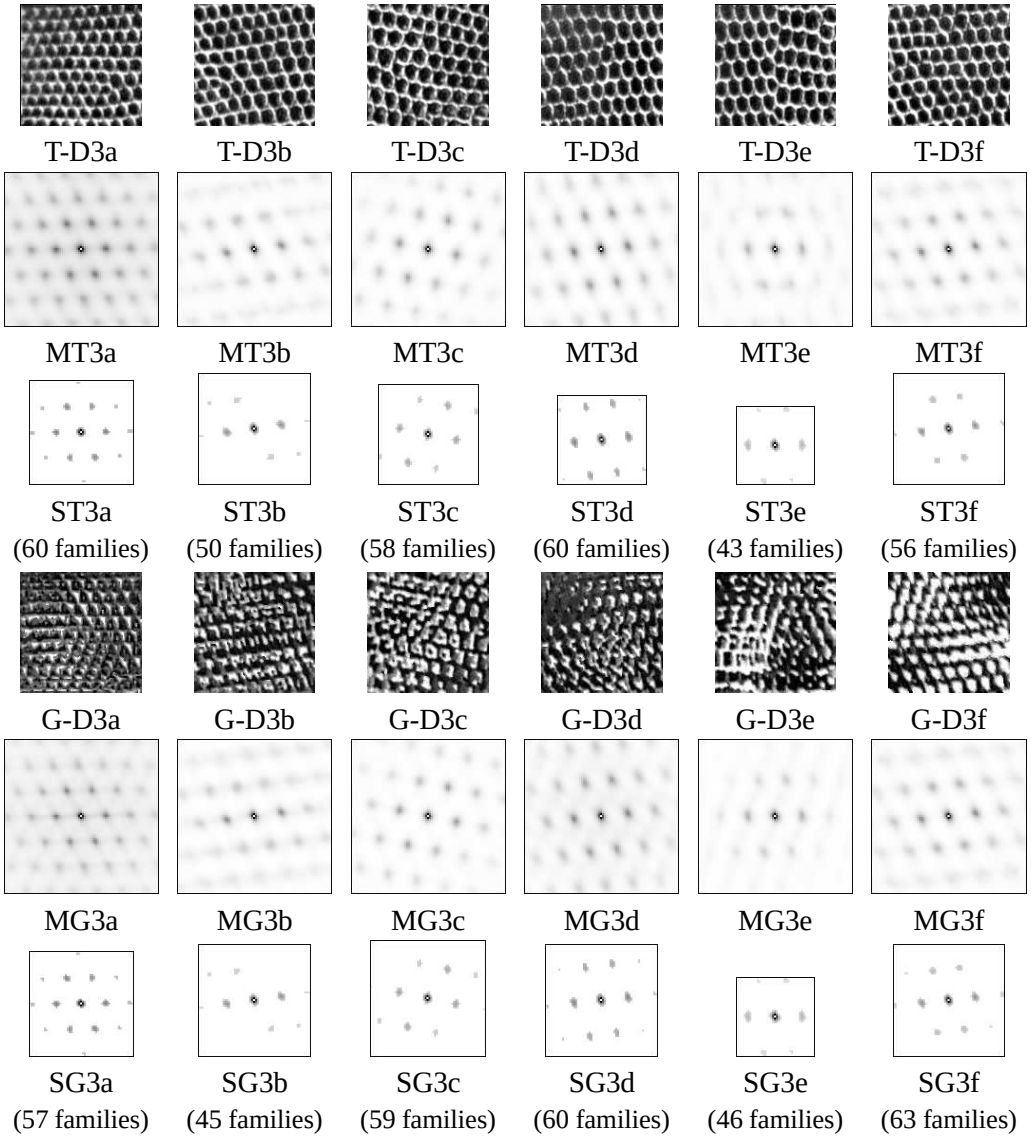


Figure 2: Texture type D3 (Reptile skin): training (T) and generated (G) samples, their interaction maps (MT, MG), and recovered structures (ST, SG) with learnt number of the families of pixel pairs.

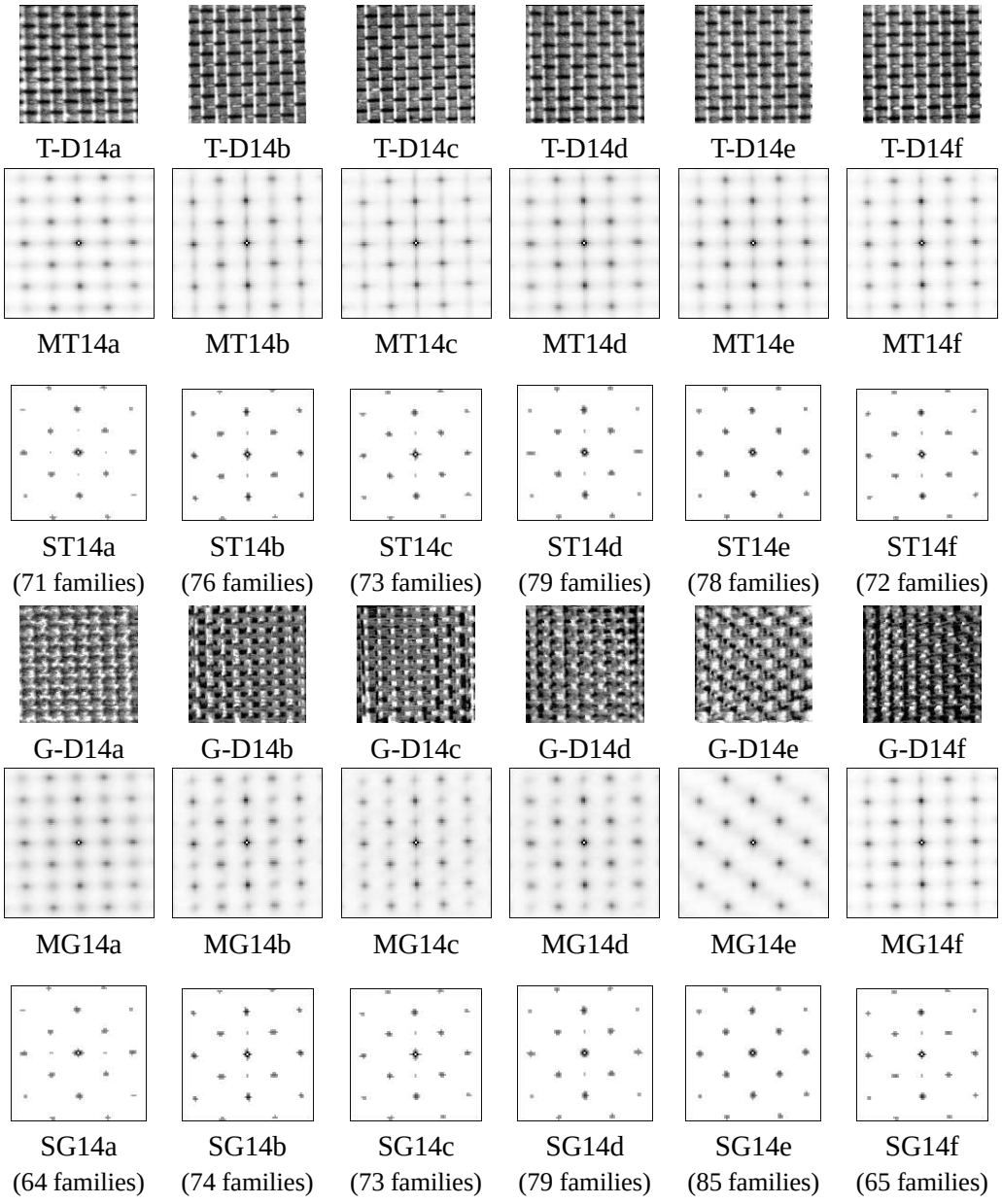


Figure 3: : Texture type D14 (Woven aluminum wire): training (T) and generated (G) samples, their interaction maps (MT, MG), and recovered structures (ST, SG) with learnt number of the families of pixel pairs.

geometric forms and numbers of the chosen clique families. Therefore, they all represent the same texture type to within the chosen Gibbs model.

Figure 4 presents two more texture samples 256×256 simulated by the CSA with the model parameters learnt for the samples T-D3a and T-D14a in Figures 2 and 3. Once again, the interaction maps and recovered structures for these larger generated samples have almost no differences relative to the ones for the smaller samples in Figures 2 and 3. In these experiments the control parameters $c_0 = 0.$, $c_1 = 1.$, $c_2 = 0.001$, and 200 macrosteps of the CSA were used to generate either sample in Figures 2 - 4.

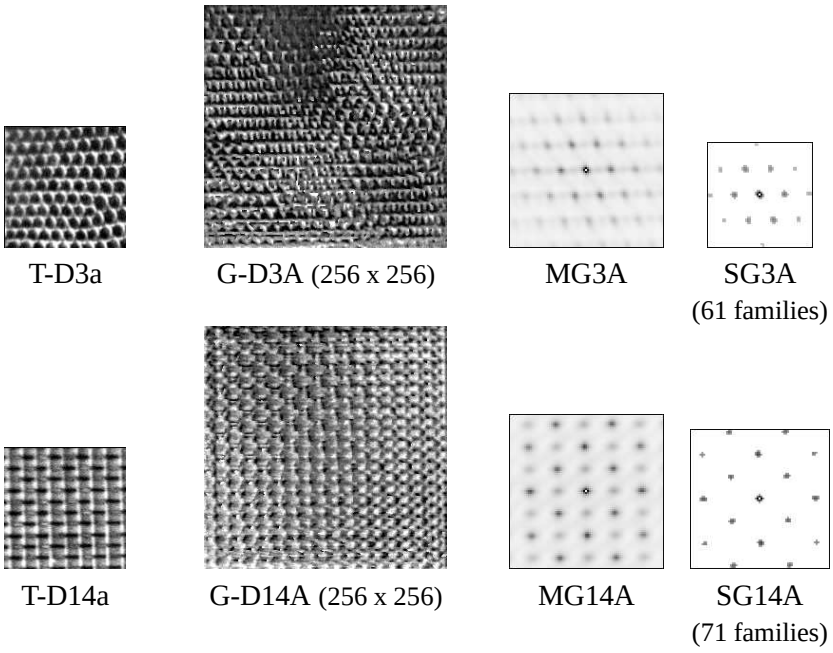


Figure 4: Training 128×128 and generated samples 256×256 of the textures D3 and D14, their interaction maps, and recovered structures for the generated samples.

It is easily seen that the simulated samples, in spite of their rather limited visual similarity to the natural ones, reflect basic spatial structure of the training patterns. Notice that the model 3 is simplified by taking into account only of the gray level differences. Because the training GLDHs possess an approximate mirror symmetry and hence are almost invariant to inversion of the image gray ranges $q \rightarrow q_{\max} - q$, the simulated images demonstrate

continuous transitions between the initial (training) and inverted image representations. To avoid such changes, a more general image model exploiting the pairwise gray level co-occurrences should be involved. But, because visually the generated samples differ notably from the natural training ones both these textures D3 and D14 are unlikely to be considered as the stochastic ones.

Figures 5 and 6 present 36 samples of other natural textures from BRODATZ [1966], used as the training ones, and the corresponding simulated samples obtained by the CSA. These experiments show that some these textures, say, D4, D5, D9, D29, D50, D57, D68, D69, D76, D77, D79, D80, D92, D93, really belong to the class of stochastic textures. In other words, the natural and simulated patterns possess both the good visual resemblance and the high proximity between their GLH and GLDHs chosen for the Gibbs model (3). But generally, such a proximity does not assure the visual similarity which is either limited (as for the textures D24, D65, D74, d82, D83, D84, D105) or nonexistent at all. In the latter case, the texture either is built from some regular texels containing more than a single pixel (D6, D11, D17, D20, D34, D55, D82, D85, D101) or the training sample has substantial local or global inhomogeneities (D23, D36, D66, D75, D95, D103). Therefore, one should discriminate between the homogeneous stochastic textures and all the other types only by such simulation experiments.

Of course, the GLDH-based interaction maps allow to check if there exist most characteristic interactions to represent the given texture type only if the texture is spatially homogeneous, or translation invariant in terms of the local conditional probabilities of the signals, and if the sample size is sufficiently large to get consistent estimates of them (as shown in Section 6). The interaction map cannot reveal structural features perceived easily by human vision if the local signal configurations are not sufficiently homogeneous over the image. In such a case, due to averaging of different non-homogeneous interactions in the GLDHs, the resulting interaction maps and learnt interaction structures do not reflect essential local features of the interactions. Also, a proper choice of the thresholds to reveal the characteristic interaction structures of different textures is not obvious and needs further theoretical justification.

In spite of these drawbacks, the proposed technique for recovering the interaction structure holds much promise in texture simulating and retrieving (GIMEL'FARB [1996a, 1996b, 1996c], JAIN and GIMEL'FARB [1995], GIMEL'FARB and JAIN [1996]). On a basis of this technique, the feature-based interaction maps, derived from the extended GLDHs, have been recently introduced by CHETVERIKOV and HARALICK ([1995]) to analyze such fea-

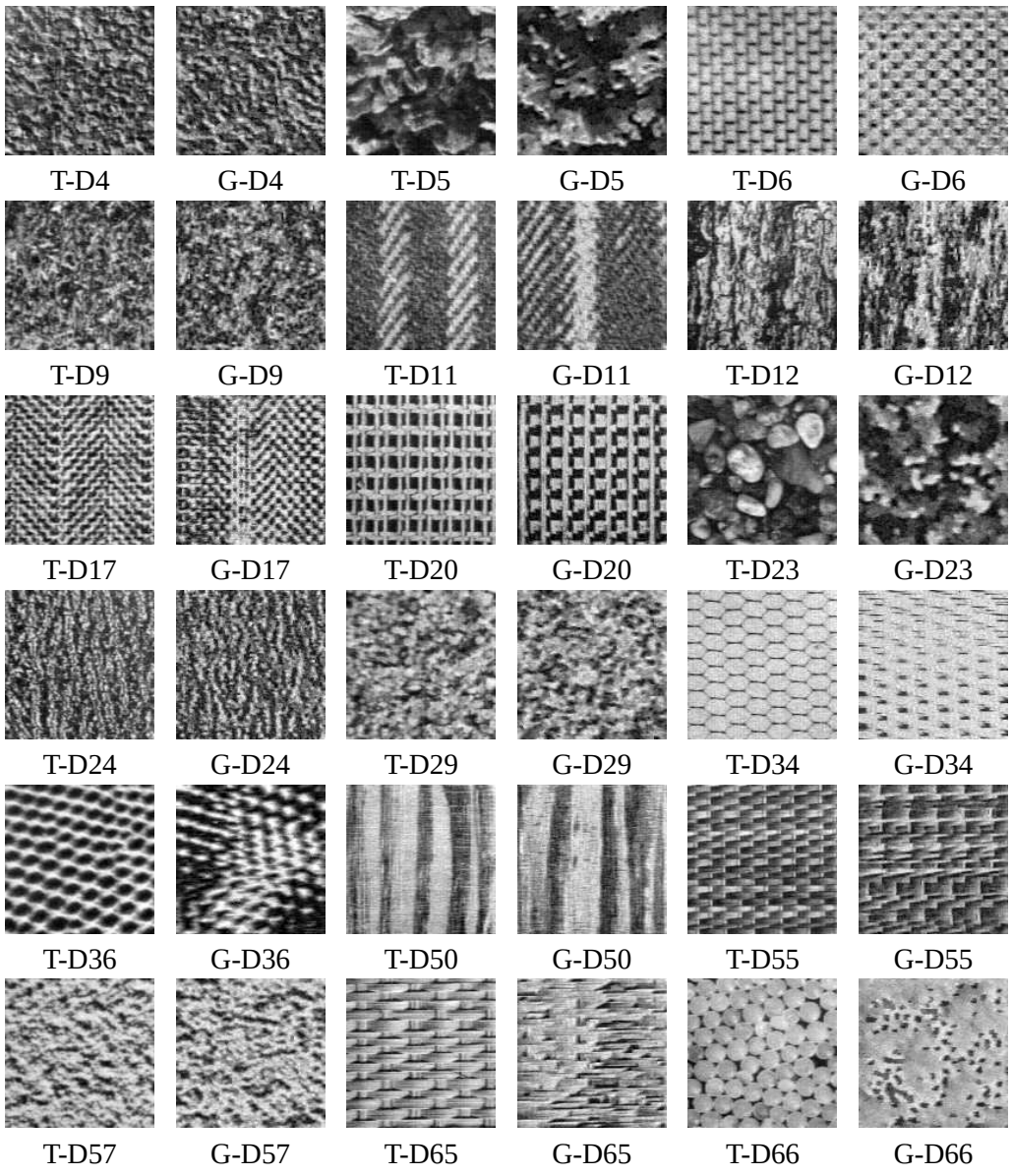


Figure 5: Training (T) and simulated (G) 128×128 samples of different natural textures.

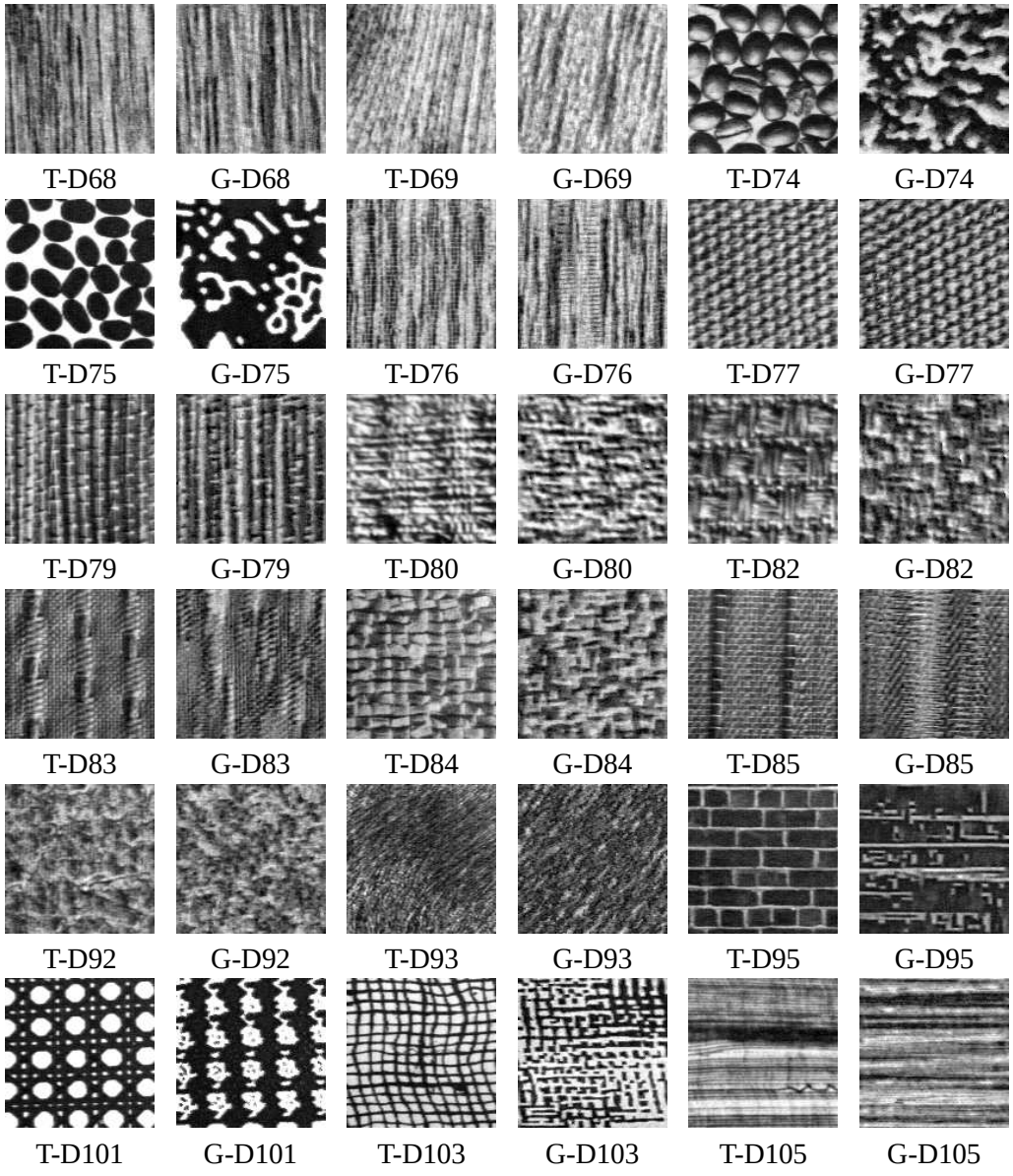


Figure 6: Training (T) and simulated (G) 128×128 samples of different natural textures.

tures of the image textures as symmetry, anisotropy, and regularity. In particular, if there exist internal symmetries in the texture patterns then the chosen structures or, at least, the initial interaction maps reveals them. For instance, let us rotate or mirror the interaction structures in Figure 4 with respect to the starting one and compute, for each the rotation or mirror angle, a relative number of the matching clique families. By a due thresholding, one can easily obtain three characteristic rotation angles and the same mirror ones for the structure SG3A or a single rotation and four mirror angles for the structure SG14A. It is such structural matching that was successfully used for retrieving the textures from an image data base in (JAIN and GIMEL'FARB [1995], GIMEL'FARB and JAIN [1996]).

Also, isolated clusters of the long-range interactions that possess the rotation or mirror symmetries and almost the same inter-cluster distances count in favour of the texture granularity. Textures with a marked lineation have corresponding "lines" of the clique families in the chosen structures, whereas the randomness leads to connected "blobs" of the chosen families around the origin. But, a quantitative description of such features needs more theoretical and experimental efforts in spite of a possibility to link these features, in principle, with the interaction maps and structures.

7.2 Supervised texture segmentation

The conditional model of region maps given a grayscale image (14) allows for a supervised segmentation of the piecewise-homogeneous textures. It should be noted that the simplified variant of the model, with the zero-valued inter-region potentials and with the resulting potential centering (19), only describes the intra-region interactions, that is, the interactions between the region labels and gray levels only within the corresponding homogeneous regions in the image. Usually such interactions are quite similar in the training pairs "a grayscale image – a region map" used for learning the model parameters and in the given test samples to be segmented. But, the inter-region interactions may differ significantly because of quite distinct region sizes and arrangements. Moreover, the total size of the subregion where the cliques of a given family cross the region borders is usually very small in comparison with the region itself. Thus, the collected GLD/RLCHs may inadequately describe the true marginal probabilities in the training pair so that the test image and its segmentation map, simulated under the general non-simplified model (14), may possess very different inter-region statistics. Therefore, in the experiments below, the characteristic interaction structure and potentials are separately learnt under the conditions (15) and under the conditions (19) from the GL/RLH and GLD/RLCHs for the same training pair. Then, the two-stage procedure is applied to a given test image to be segmented. At the first stage, the simplified model (14) – (19) is used to obtain by the CSA an initial segmentation map

for the given image, starting from a sample of the IRF of the region labels. At the second stage, the general model (14) – (15) is applied to form by the CSA a final map, starting from the obtained initial one.

7.2.1 Artificial collages of natural textures

Figure 7 shows the 5-region collage (256×256) of natural textures from BRODATZ [1966] to be segmented using the conditional model (14) and its ideal region map. The training collage (256×256) and its region map are presented in Figure 8. This training pair is formed by a random arrangement of 5 patches (64×64) of these texture types. There are 4 randomly chosen positions for the patch of type 0 and 3 such positions for the patch of each other type $1, \ldots, 4$.

The initial segmentation map, shown in Figure 9*a*, is obtained after 300 macrosteps of the CSA with the control parameters $c_0 = 0;\ c_1 = 1;\ c_2 = 0.001$. Figure 9*b*, demonstrates the final segmentation map obtained also after 300 macrosteps of the CSA with the same parameters. The final map is sufficiently close to the ideal one in spite of small deviations of the region borders. The deviations are mostly due to local similarities between small patches of these textures. Also, the inter-region interactions in the training sample slightly differ from the same interactions in the test collage because of the distinct region borders and arrangement. Figure 9*c* displays the segmentation errors. Here, black pixels indicate all the positions with different labels in the ideal and segmentation maps. The relative error rate is 3.08%.

Figure 10 demonstrates separately the texture regions obtained by the final segmentation. It is evident that the found regions are equivalent to the ideal ones with respect to the intra-region texture homogeneity.

Figures 11 - 18 present results of segmenting 4-region collages (256×256) of different homogeneous, weakly homogeneous, and almost inhomogeneous textures from BRODATZ [1966]. All the training collages have the same region map with 4 regions $0, \ldots, 3$ formed each by a random arrangement of 4 square patches (64×64). Thus, each the training region contains 16384 pixels. The training collage cuts the corresponding patches from the initial textures (256×256) ranked by their serial numbers in BRODATZ [1966]: say, D3, D4, D5, and D9 in Figure 11; D11, D12, D17, D20 in Figure 12, etc. Also, all the test collages share a single region map generated by using the Markov/Gibbs map model (6) with parameters learnt from the training map. The regions 0,1,2,3 in the test map have different areas: 23382, 17502, 17604, and 7048 pixels, respectively. The test collage also cuts the

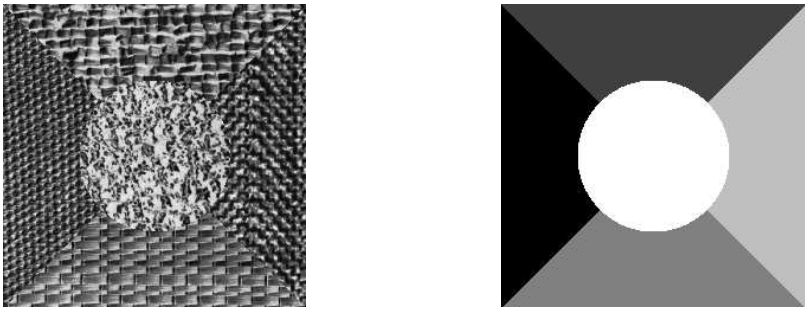


Figure 7: Collage 256×256 to be segmented and its ideal region map.

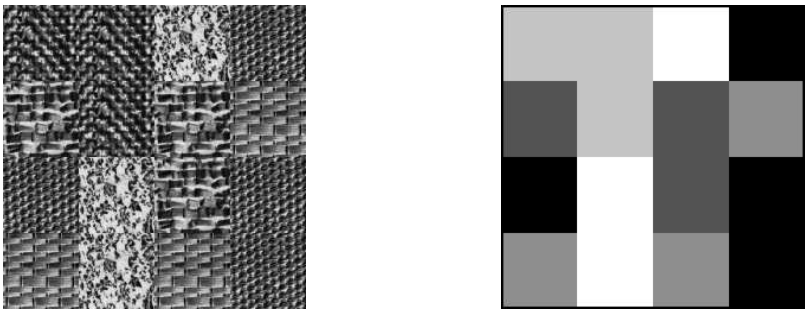


Figure 8: Training collage 256×256 and its region map.

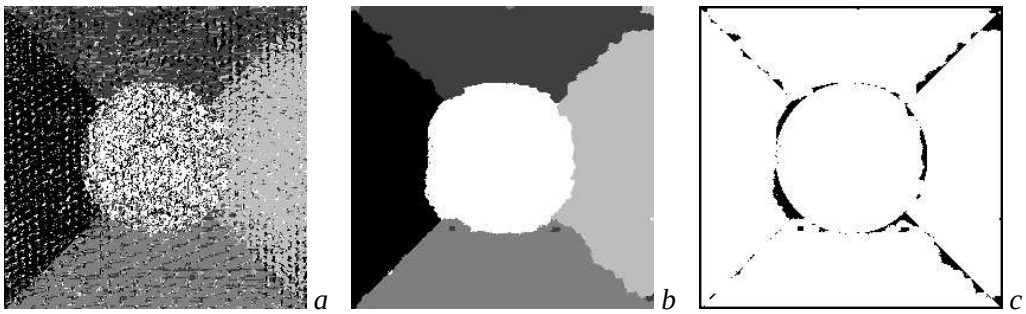


Figure 9: Initial and final segmentation maps and positions of the segmentation errors.

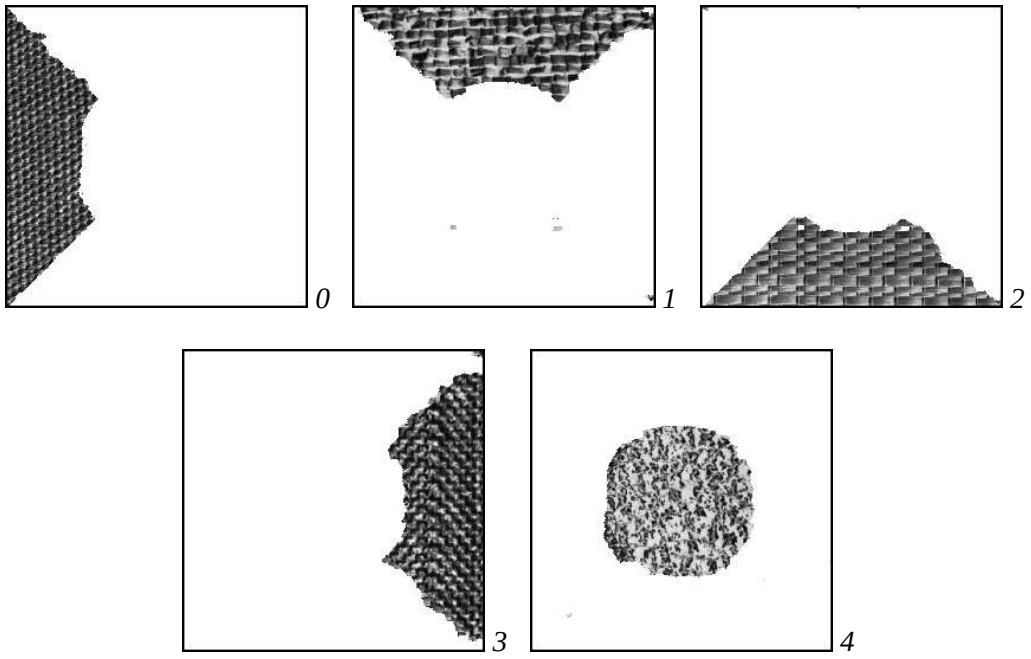


Figure 10: Texture regions after the segmentation.

corresponding patches from the same initial textures but, because both the region maps are distinct, it contains much different texture patches than the training one.

Figures 11 - 18 display the training and test grayscale collages, their region maps, and initial and final segmentation maps for the training and for the test collage. Both these latter maps are obtained using 300 CSA macrosteps and the learnt from the training pair parameters of the model. Table 3 gives total error rates for these maps in terms of relative numbers of the misclassified pixels (that is, of the pixels having different labels in the ideal and obtained maps).

Table 3: Relative segmentation errors (in percents)

Fig.	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Text-ures	D3,4, 5,9	D11,12, 17,20	D23,24, 29,34	D50,55, 57,65	D66,68, 69,74	D75,76, 77,79	D83,84, 85,92	D93,95, 101,103
Trai-ning	2.46	0.31	3.44	0.01	0.05	0.02	2.67	0.42
Test	25.50	20.98	34.43	19.20	19.16	21.01	19.67	16.78

The above segmentation results indicate that the training samples have notable distinctions from the test ones as regarding the statistics of the pairwise interactions. It is also obvious from Figure 19 presenting the results of segmenting one more test collage with the textures D23, D24, D29, and D34. Its region map is similar to the training one but has an "complement" region labeling (namely, the region number $3 - k$ instead of the number $k = 0, \ldots, 3$ in the training map). In this case, the total error rate is significantly lower than in Figure 13 (7.20% vs. 34.43 %). Thus, the closer the inter-region interactions in the test images to the training sample, the lower the segmentation error.

It is worth noting that borders between the texture regions present main difficulties to the proposed segmentation because it is the inter-region statistics (and, therefore, the potentials) that most likely differ in the test images relative to the training pair. The more the textures to be separated in the test image, the less definite the reconstruction of their borders. This is especially true for our experiments because (*i*) the simplified model (14) with the fixed interaction structure for all the textures is used and (*ii*) the characteristic interaction structure is recovered by a straightforward thresholding of the relative partial energies in the interaction map with no account of a resulting discrimination between the individual texture types. Thus, there exist potentialities to amplify the obtained experimental results to

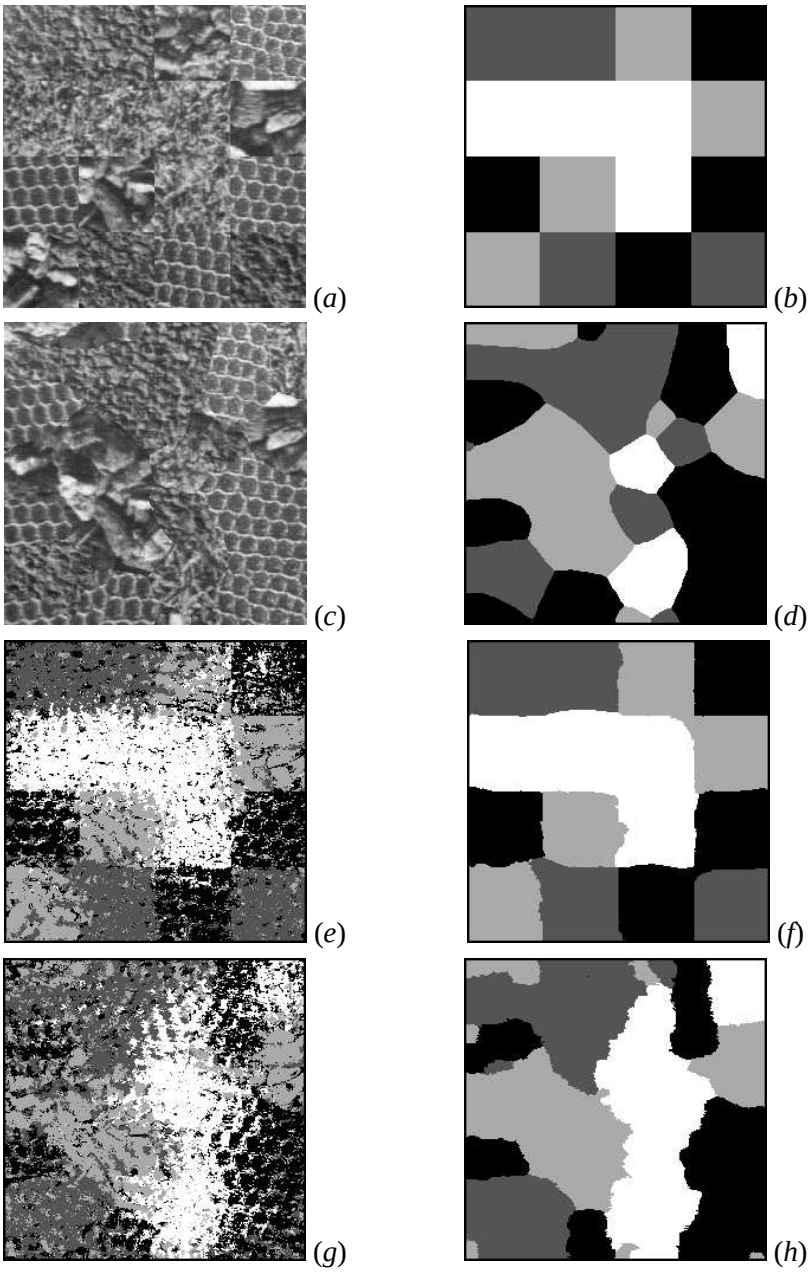


Figure 11: Segmentation of the 4-region collage of the textures D3, D4, D5, D9 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

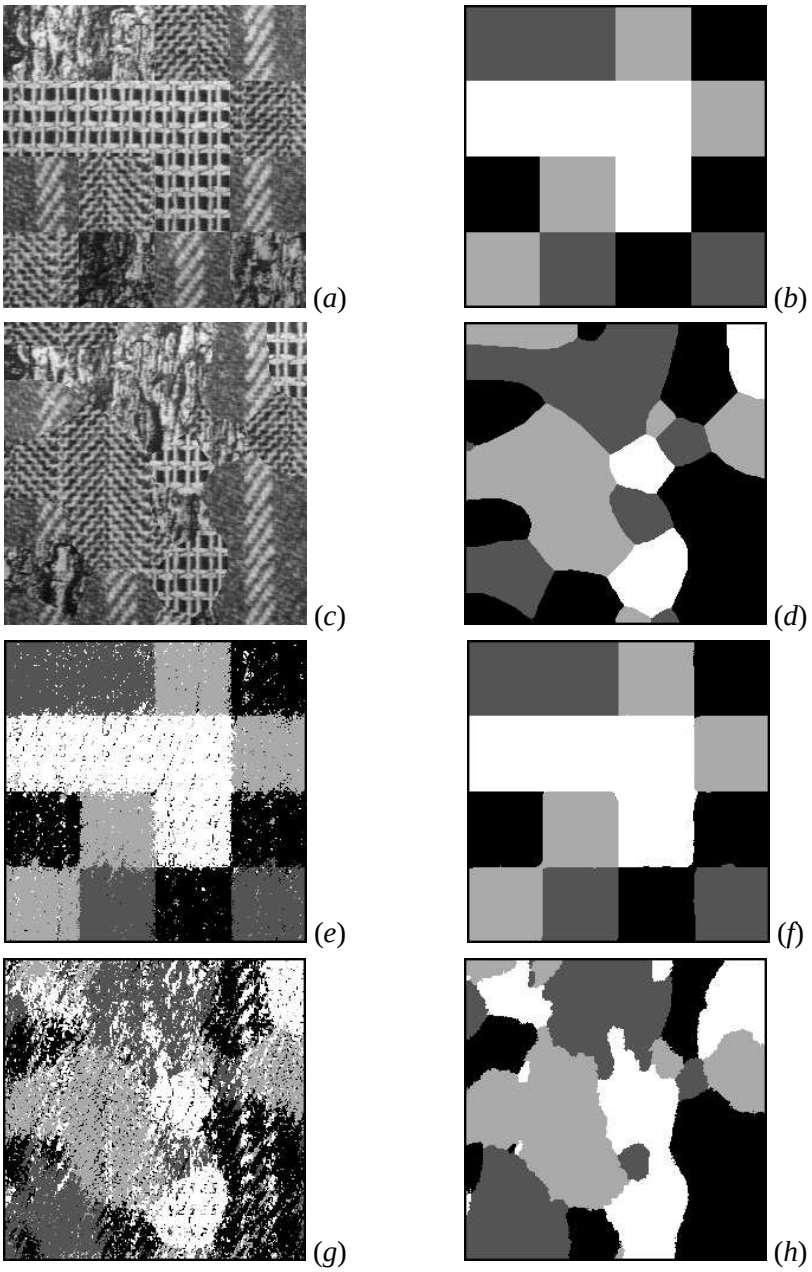


Figure 12: Segmentation of the 4-region collage of the textures D11, D12, D17, D20 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

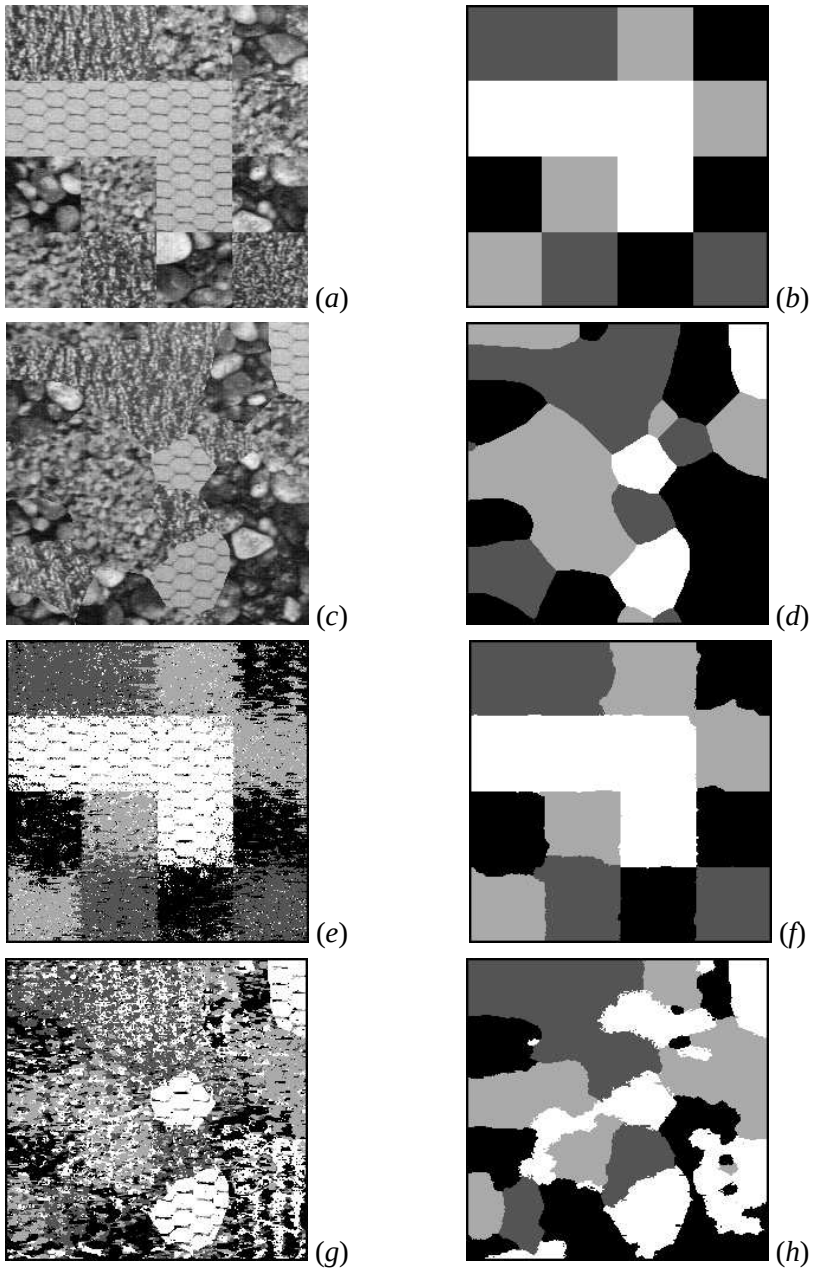


Figure 13: Segmentation of the 4-region collage of the textures D23, D24, D29, D34 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

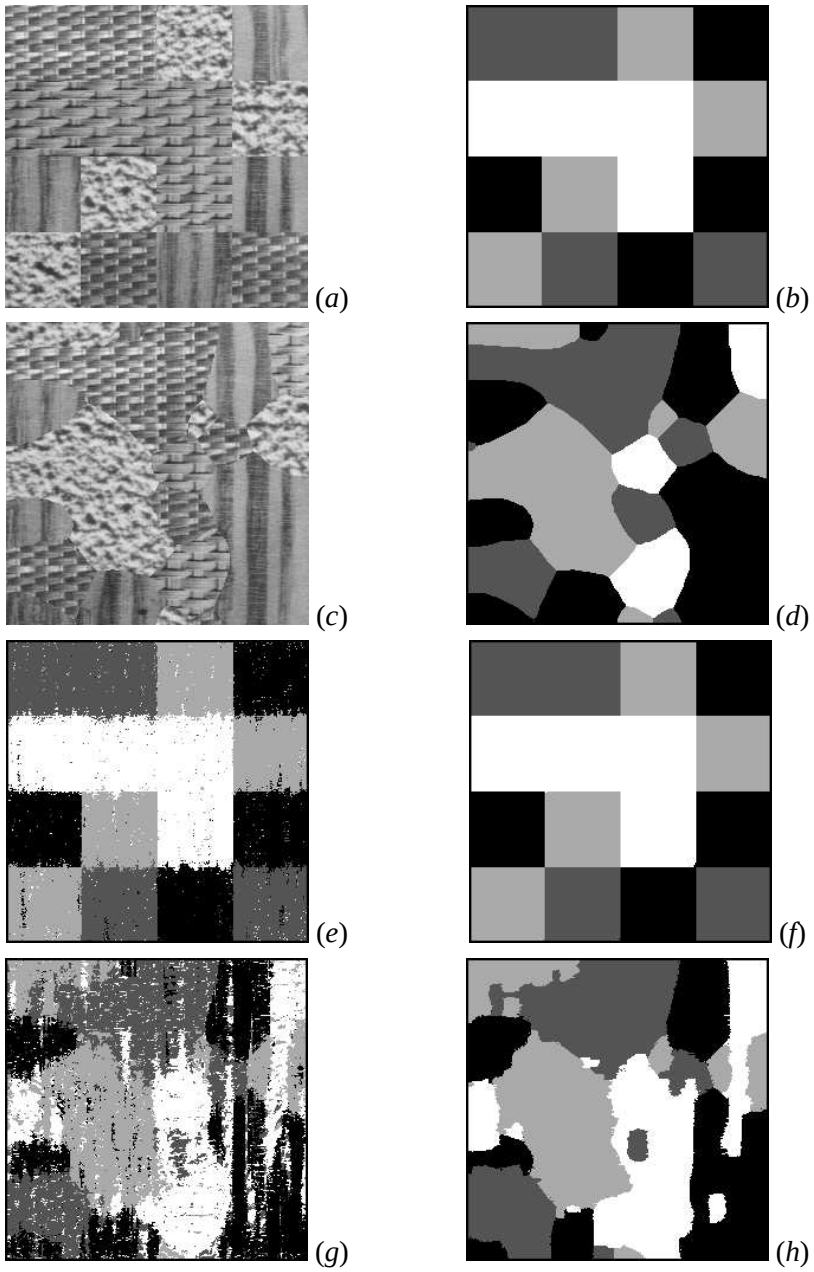


Figure 14: Segmentation of the 4-region collage of the textures D50, D55, D57, D65 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

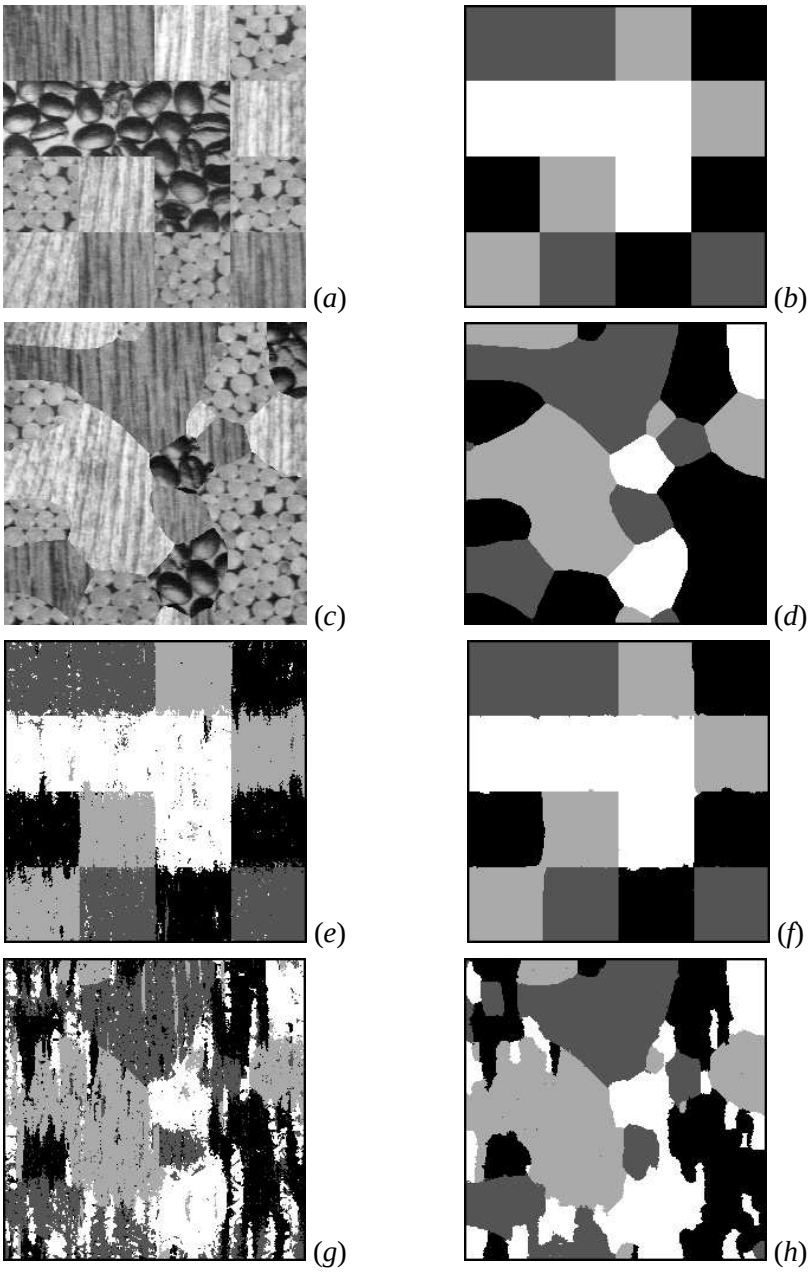


Figure 15: Segmentation of the 4-region collage of the textures D66, D68, D69, D74 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

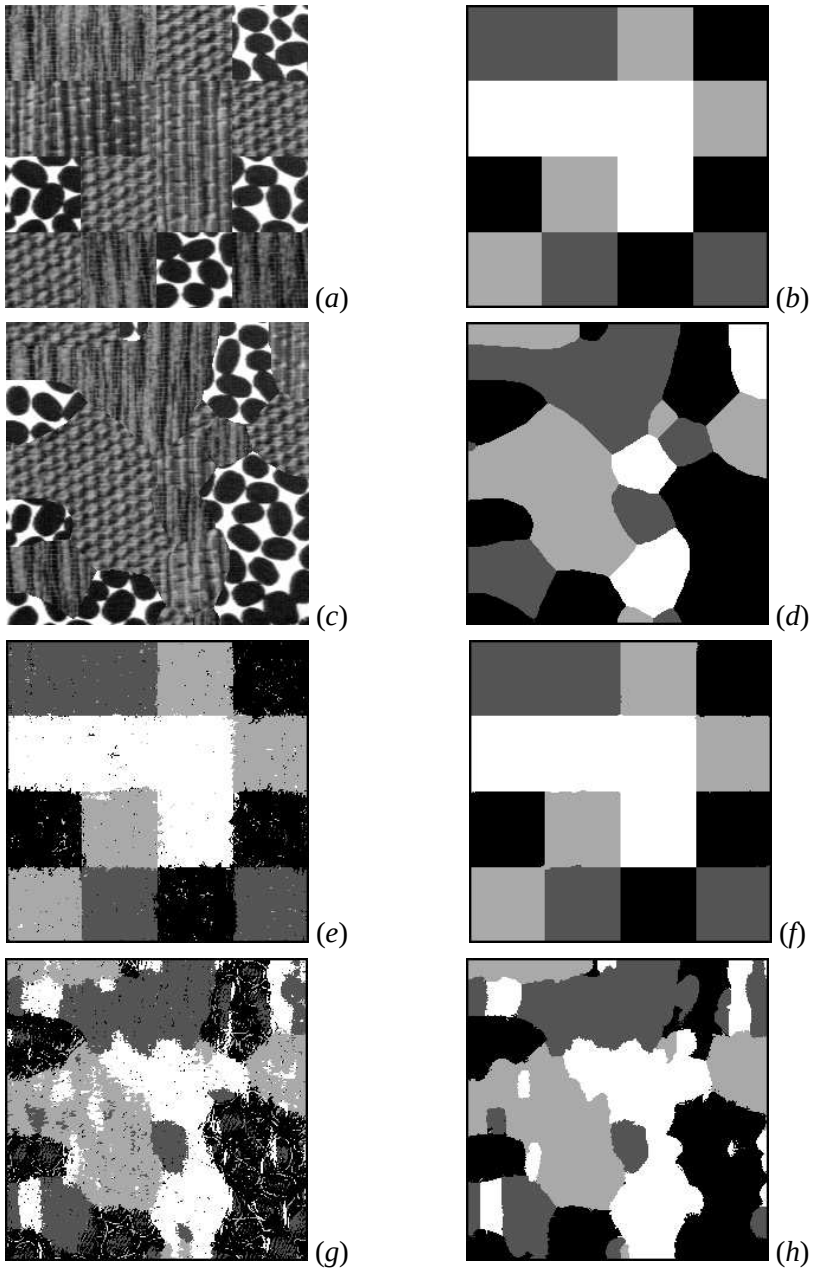


Figure 16: Segmentation of the 4-region collage of the textures D75, D76, D77, D79 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

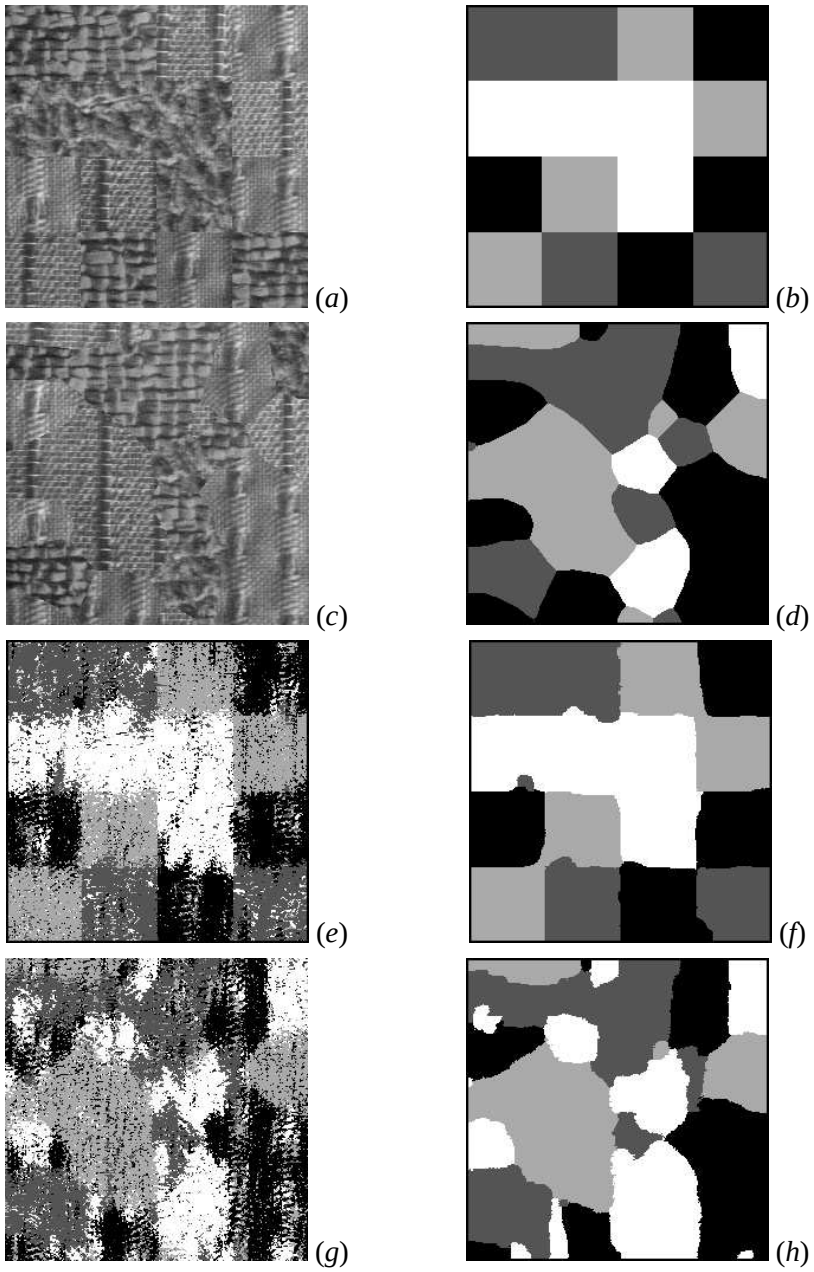


Figure 17: Segmentation of the 4-region collage of the textures D83, D84, D85, D92 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

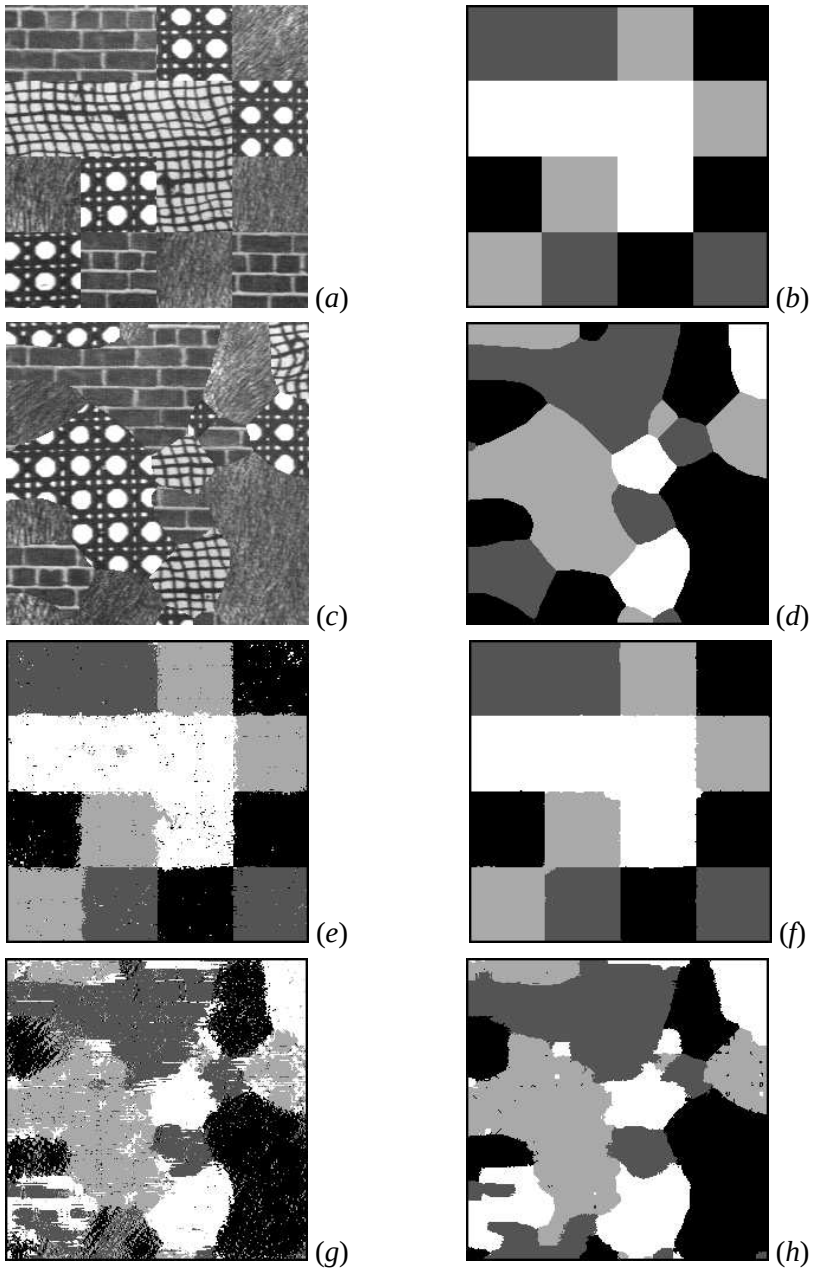


Figure 18: Segmentation of the 4-region collage of the textures D93, D95, D101, D103 (the training collage (*a*) and its region map (*b*), the test collage (*c*) and its ideal map (*d*), the initial (*e*) and final (*f*) segmentation maps for the training collage, and the initial (*g*) and final (*h*) segmentation maps for the test collage).

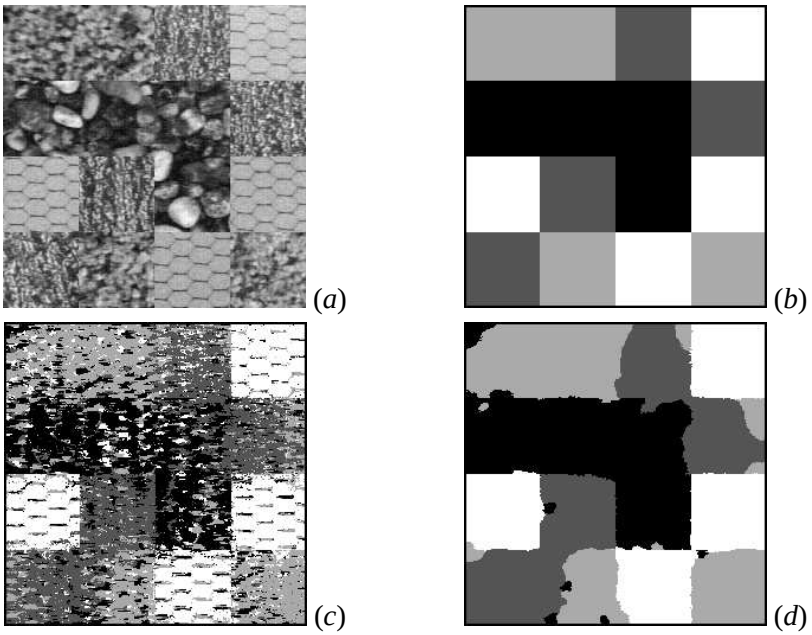


Figure 19: Segmentation of the 4-region collage of the textures D23, D24, D29, D34 (the test collage (*a*), its ideal map (*b*), and the initial (*c*) and final (*d*) segmentation maps; the training pair is the same as in Figure 13).

be investigated in future.

Figures 20, 21, and 22 demonstrate the 16-region test collage (512×512) of various homogeneous, weakly homogeneous, and almost inhomogeneous textures from BRODATZ [1966] and its ideal region map, the training pair (256×256) with randomly arranged 16 patches (64×64, one patch per texture type), and the resulting initial and final segmentation maps. The initial map in Figure 22*a* is obtained by a two-stage CSA: (*i*) 1000 macrosteps with the control parameters $c_0 = 0.,\ c_1 = 10.,\ c_2 = 1.$ at the first stage and (*ii*) additional 300 CSA-macrosteps with the control parameters $c_0 = 0.,\ c_1 = 1.,\ c_2 = 0.001.$ at the second stage. The control parameters are changed at the second stage as to accelerate the convergence of the joint GLD/RLCHs collected for the test image and current segmentation map to the desired training histograms. The total chi-square distance between these intra-region GLD/RLCHs is, respectively, reduced from $172,000$ to $32,200$ at the first stage and from $32,200$ to $12,400$ at the second stage. The final segmentation map is obtained by a similar two-stage CSA: (*i*) 300 macrosteps with the control parameters $c_0 = 0.,\ c_1 = 10.,\ c_2 = 1.$ to reduce the total chi-square distance between the intra- and inter-region GLD/RLCHs from $1,603,000$ to $67,400$ at the first stage and (*ii*) subsequent 300 CSA-macrosteps with the control parameters $c_0 = 0.,\ c_1 = 1.,\ c_2 = 0.001.$ to reduce this distance from $67,400$ to $52,100$ at the second stage.

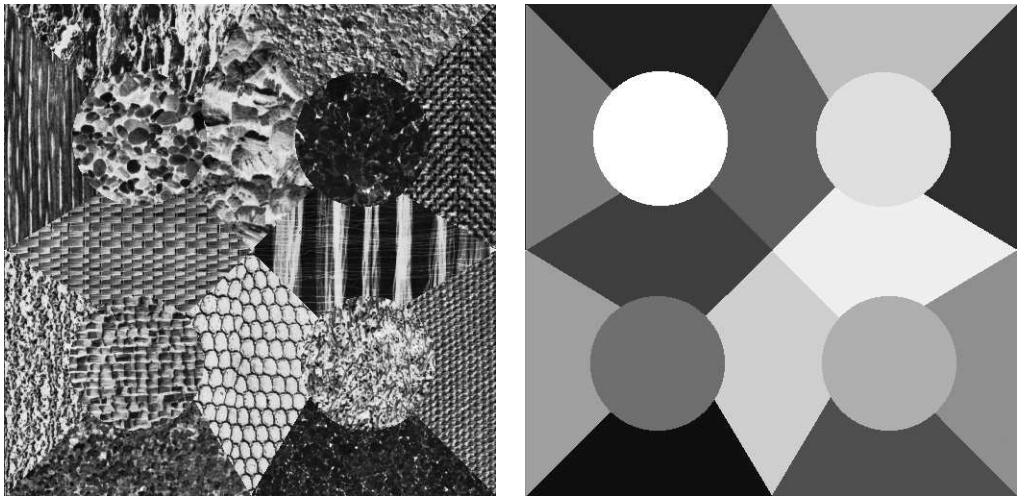
Figures 24 – 27 shows 16 texture regions found by the final segmentation. The segmentation errors, that is the positions with different labels in the ideal region map and final segmentation map, are depicted by black pixels in Figure 23*a*. Figure 23*b* shows the texture pieces which cause the segmentation errors. The total segmentation error (that is difference between the obtained and ideal region maps) is about 13.82% (32239 pixels with different labeling among 262144 pixels in the lattice). Table 4 gives absolute and relative errors for each region in the test image. It is easily seen that most differences between the ideal and segmentation maps are due to similarities between subparts of different weakly homogeneous textures. The texture inhomogeneities result in significant distinctions between the inter- and intra-region statistics in the training and segmented images. But, visually, the found textures are sufficiently homogeneous over the regions obtained by segmenting.

7.2.2 Natural multi-textured images

Figures 28 and 29 show results of segmenting a natural SAR image of the Earth's surface. In this case, the training map contains small and mostly disjoint patches of the desired

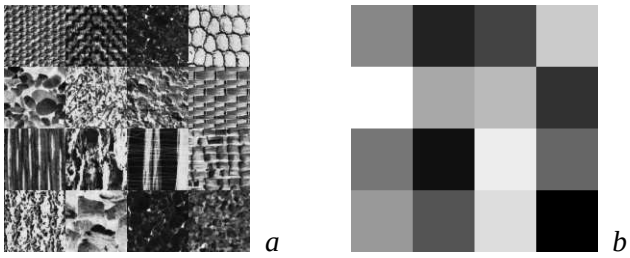
Table 4: Absolute and relative errors in the final 16-region segmentation map.

Region	Size in the ideal map	Errors	In %
0	15057	564	3.7
1	15241	3144	20.6
2	15321	634	4.1
3	19814	3227	16.3
4	15127	424	2.8
5	19503	7003	35.9
6	15440	1539	10.0
7	15311	716	4.7
8	14696	298	2.0
9	15007	1133	7.5
10	15440	1121	7.3
11	15374	1053	6.8
12	20281	4594	22.7
13	15440	810	5.2
14	19652	5682	28.9
15	15440	4297	27.8



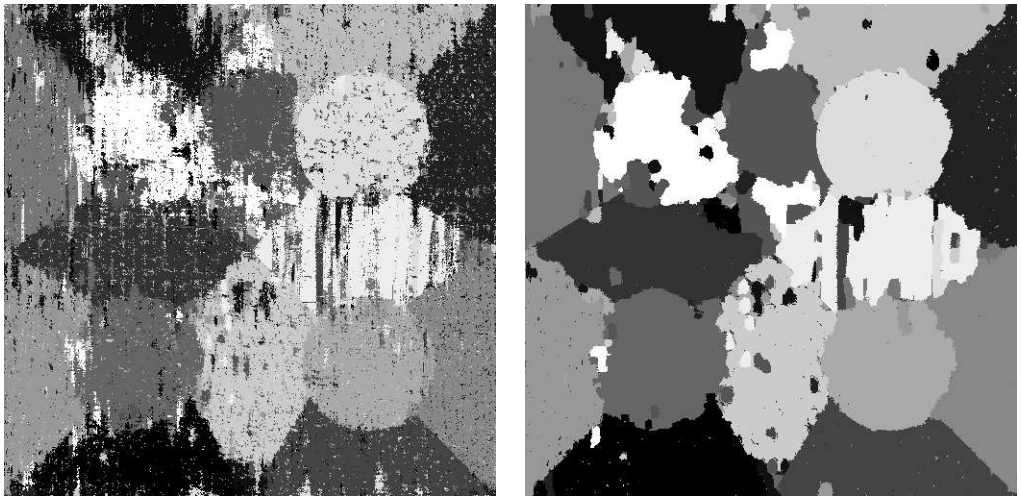
a *b*

Figure 20: 16-region texture collage 512×512 (*a*) and its ideal region map (*b*).



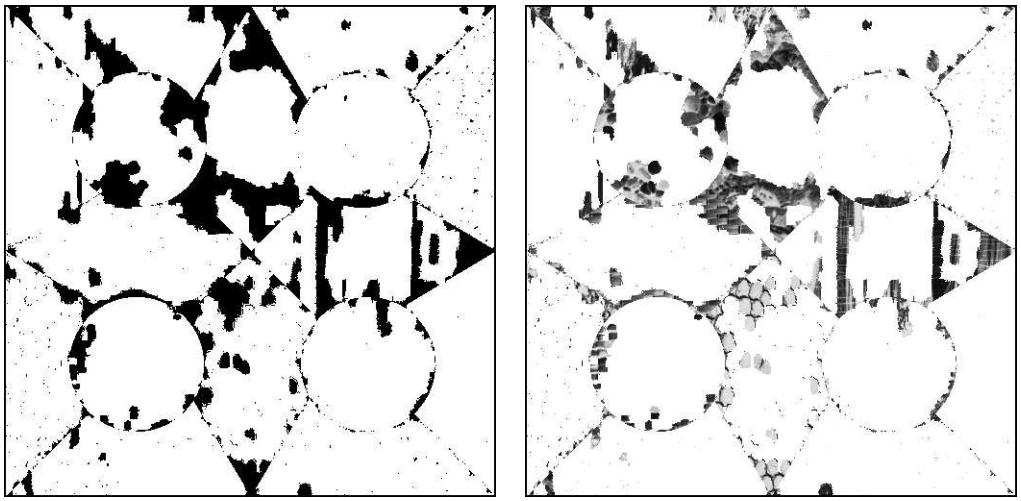
a *b*

Figure 21: 16-region training collage 256×256 (*a*) and its region map (*b*).



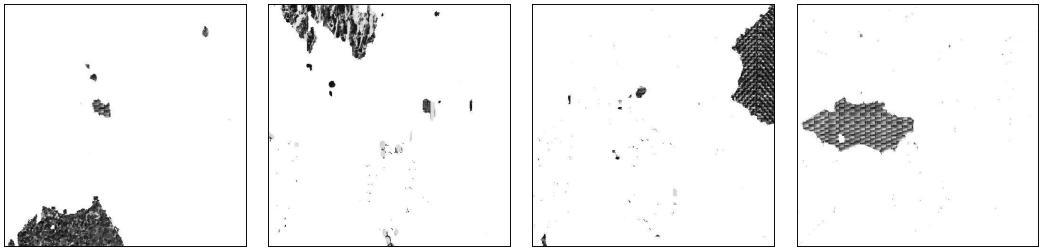
a *b*

Figure 22: Initial(*a*) and final (*b*) 16-region segmentation map.



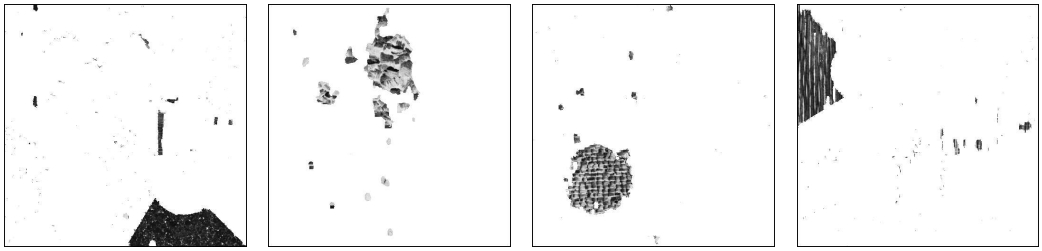
a 2mm*b*

Figure 23: Segmentation errors relative to the ideal 16-region map (*a*) and corresponding textures in the test image (*b*).



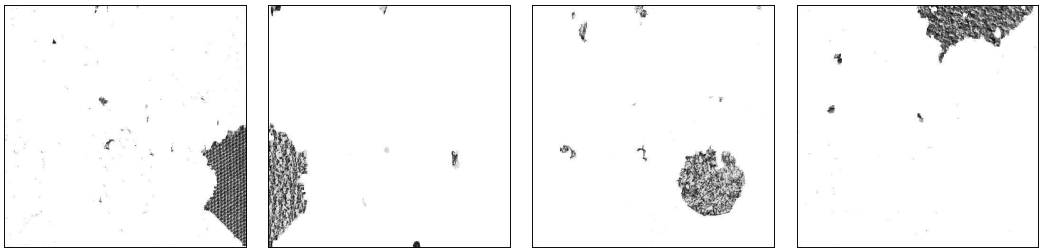
Region *0* Region *1* Region *2* Region *3*

Figure 24: Texture regions $0 \ldots 3$ after the segmentation (on a reduced scale).



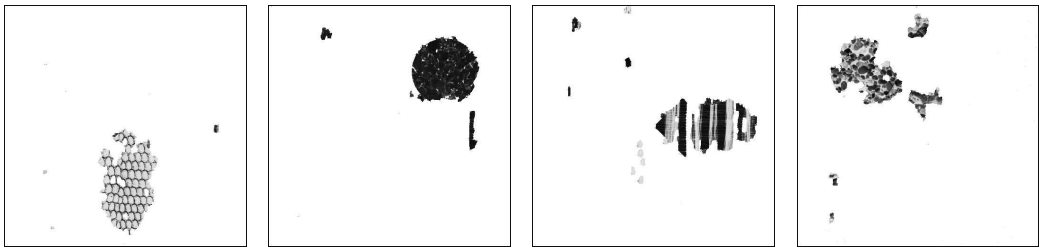
Region *4* Region *5* Region *6* Region *7*

Figure 25: Texture regions $4 \ldots 7$ after the segmentation (on a reduced scale).



Region *8* Region *9* Region *10* Region *11*

Figure 26: Texture regions $8 \ldots 11$ after the segmentation (on a reduced scale).



Region *12* Region *13* Region *14* Region *15*

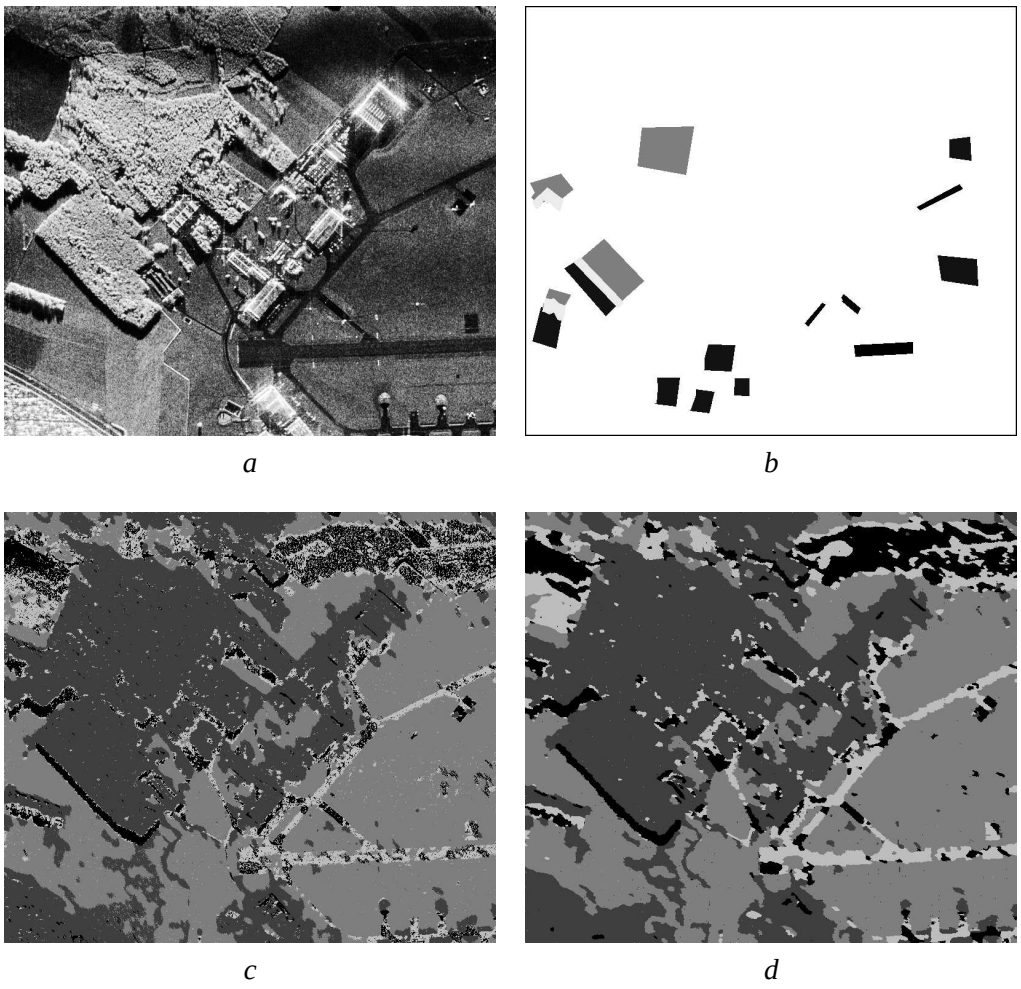
Figure 27: Texture regions $12 \ldots 15$ after the segmentation (on a reduced scale).

4 textured regions chosen visually in the initial image. The training textured objects include radio-shadows (the region 0), vegetation areas (the region 1), grass fields (the region 2), and concrete ways (the region 3). Notice that the buildings are absent in the training data. Thus, they are mostly included into the region 1 having most similar GLH and close-range GLDHs. The final segmentation map gives rather correct discrimination between the regions 1 and 2 but presents some errors as regarding the discrimination between the radio-shadows and concrete ways. These latter texture types possess a close similarity in terms of their GLHs and close-range GLDHs.

7.3 Concluding remarks

These theoretical and experimental results allow to conclude that the proposed Markov and non-Markov Gibbs models with multiple pairwise pixel interactions hold much promise in texture simulation and segmentation. They possess the following features which facilitate their use in practice:

- The models are based on the explicit assumptions about the grayscale images and region maps under consideration.
- The joint model of piecewise-homogeneous images and region maps generalizes in a straightforward way the models of the homogeneous images and of the region maps so that these latter ones can be considered as particular cases of the piecewise-homogeneous model.
- The conditional models of the piecewise-homogeneous grayscale images, given a region map, and of the region maps, given a grayscale image, are obtained easily from the joint model by fixing either the region map or the grayscale image.



a *b*

c *d*

Figure 28: Initial SAR image (*a*), map of the training patches (*b*), initial (*c*) and final (*d*) segmentation maps.

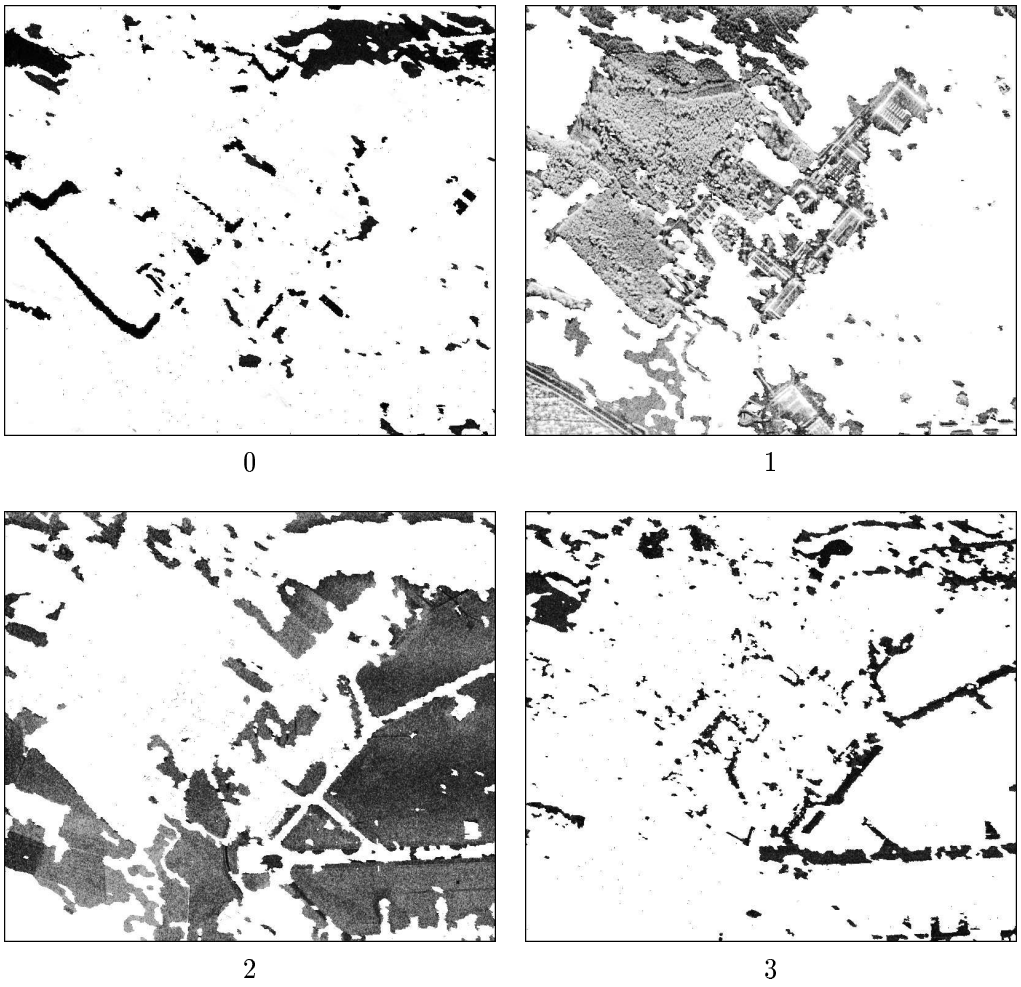


Figure 29: Texture regions $0, \ldots, 3$ in the final segmentation map.

- Generally, the conditional piecewise-homogeneous models possess various interaction structures for the regions with the same label or with the same gray level difference.
- Due to their internal symmetries, all the models exploit the like parameter learning scheme and embed both the texture simulation and segmentation into the same Bayesian decision framework.
- The learning scheme, based on the MLE of the Gibbs potentials, allows to estimate both the characteristic structure of the interactions and the potentials for the chosen structure by using, first, the analytic and, then, the stochastic approximation of the potentials.
- The conditional MLE of the potentials, provided that the training sample may rank a feasible top place in a parent population by the total Gibbs energy, allows to obtain the explicit, to scaling factors, potentials and reduce the learning only to these factors.
- The Bayesian decisions regarding the desired image or region map are approximated by generating the Markov chain of the images or maps by the controllable simulated annealing.

There are many open problems to be solved in future, in particular, an optimal search for the interacting structure, a theoretical justification of the control parameters and stopping rules to be used for the potential refinement and image simulation, unsupervised learning techniques to exclude or reduce the supervised learning of the inter-region interactions which are most different in the training and test images, unsupervised texture segmentation using specific models with multiple pairwise pixel interactions, *etc*...

Acknowledgements

It is difficult to overestimate multiple pairwise interactions, discussions, and stimulating collaboration with Dmitry Chetverikov, Wolfgang Förstner, Anil Jain, Mikhail Schlesinger, Marc Sigelle, Alexey Zalesny, and Josiane Zerubia. I am very grateful to Anil Jain for his invaluable help in improving the quality and clarity of the first theoretical and experimental results.

I would like to thank Hungarian Academy of Sciences (Hungary), Deutsche Forschungsgemeinschaft (Germany), and Ministere de l'Education Nationale, de l'Enseignement Superieur, et de la Recherche (France) for supporting this research at different times and places.

I am indebted to Dmitry Chetverikov (Computer and Automation Research Institute, Hungary) for the digitized texture fragments from BRODATZ [1966], to Alexey Zalesny (IRTCITS, Ukraine)

for the artificial 5- and 16-region texture collages, and to Martin Kirscht (University of Hannover, Germany) for the natural SAR image and its training map used in the texture simulation and segmentation experiments.

References

[1966] Abend, K.: Compound decision procedures for pattern recognition. Proc. Nat. Electron. Conf., vol. 22 (1966) 777-780

[1970] Averintsev, M. B.: On one method of describing random fields with discrete argument. Problems of Information Transmission 6:2 (1970) 100-108 [*In Russian*]

[1972] Averintsev, M. B.: Description of Markov random fields using Gibbs conditional probabilities. Probability Theory and Its Applications, XVII:1 (1972) 21-35 [*In Russian*]

[1990] Barnhart, C.L., Barnhart, R.K. (Eds): The World Book Dictionary Vol. 2 (L-Z). Chicago: World Book (1990) 2077, 2170.

[1978] Barndorff-Nielsen, O.: Information and Exponential Families in Statistical Theory. Wiley (1978)

[1974] Besag, J. E.: Spatial interaction and the statistical analysis of lattice systems. J. Royal Statistical Soc. B36 (1974) 192-236

[1966] Brodatz,P.: Textures: A Photographic Album for Artists an Designers. New York: Dover Publications (1966)

[1993] Chellappa, R., Jain, A. (Eds): Markov Random Fields: Theory and Application. Academic Press (1993)

[1987] Chetverikov, D.: On some basic concepts of texture analysis. Proc. 2nd Int. Conf. on Computer Analysis of Images and Patterns. Sept. 2-4, 1987, Wismar, GDR. Berlin: Akademie-Verlag (1987) 196-201

[1995] Chetverikov, D., Haralick, R. M.: Texture anisotropy, symmetry, regularity: recovering structure and orientation from interaction maps. Proc. 6th British Machine Vision Conf. September 11-14, 1995, Birmingham. Sheffield: Univ. of Sheffield (1995) 57-66

[1983] Cross, G. R., Jain, A. K.: Markov random field texture models. IEEE Trans. Pattern Anal. Machine Intell. 5:1 (1983) 25-39

[1984] Derin, H., Elliot, H., Cristi, R., Geman, D.: Bayes smoothing algorithm for segmentation of images modelled by Markov random fields. IEEE Trans. Pattern Anal. Machine Intell. 6:6 (1984) 707-720

[1968] Dobrushin, R. L.: Gibbs random fields for the lattice systems with pairwise interaction. Functional Analysis and Its Applications 2:4 (1968) 31-43 [*In Russian*]

[1975] Dobrushin, R. L., Pigorov, S. A.: Theory of random fields. Proc. 1975 IEEE-USSR Joint Workshop Information Theory, Dec. 15-19, 1975, Moscow, USSR. IEEE (1976) 39-49

[1989] Dubes R. C., Jain, A. K.: Random Field Models in Image Analysis. J. of Applied Statistics 16:2 (1989) 131-164

[1984] Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Machine Intell. 6:6 (1984) 721-741

[1996a] Gimel'farb, G. L.: Texture modeling by multiple pairwise pixel interactions. IEEE Trans. Pattern Anal. Machine Intell. 18:11 (1996) 1110-1114

[1996b] Gimel'farb, G. L.: Non-Markov Gibbs texture model with multiple pairwise pixel interactions. Proc. 13th IAPR Int. Conf. Pattern Recognition, vol. II, Aug.25-29, 1996, Vienna, Austria. TUWien (1996) 591-595

[1996c] Gimel'farb, G. L.: Gibbs models for Bayesian simulation and segmentation of piecewise-uniform textures. *Ibid*. 760-764

[1997] Gimel'farb, G. L., Schmidt, J., Braunmandl, A.: Gibbs fields with multiple pairwise interactions as a tool for modelling grid-based data. Proc. Int. Workshop on Process Modelling and Landform Evolution, Febr. 17-18, 1997, Bonn, Germany. Lecture Notes in Earth Sciences. Springer (1997) [*In print*]

[1996] Gimel'farb G. L., Jain, A. K.: On retrieving textured images from an image data base. Pattern Recognition 29:9 (1996) 1461-1483

[1979] Haralick, R. M.: Statistical and Structural Approaches to Textures. Proc. IEEE 67:8 (1979) 786-804

[1992] Haralick, R. M., Shapiro,L. G.: Computer and Robot Vision. Vol. 1. Reading: Addison-Wesley Publ. Co. (1992) 453-507

[1980] Hassner, M., Sklansky, J.: The use of Markov random fields as models of textures. Computer Graphics Image Processing 12:4 (1980) 357-370

[1971] Isihara, A.: Statistical Physics. Academic Press (1971).

[1989] Jacobsen, M.: Existence and unicity of MLE in discrete exponential family distributions. Scandinav. J. Statistics 16 (1989) 335-349

[1995] Jain A. K., Gimel'farb, G.: Retrieving textured images from an image data base. Proc. of the 9th Scandinavian Conf. on Image Analysis. June 6-9, 1995, Uppsala, Sweden. Uppsala: SSAIA (1995) 441-448

[1981] Julesz, B.: Textons, the elements of texture perception, and their interactions. Nature 290 (1981) 91-97

[1986] Kashyap, R. L.: Image models. Handbook on Pattern Recognition and Image Processing (T. Y. Young, K.-S. Fu, Eds). Academic Press (1986) 247-279

[1983] Kashyap, R. L., Chellappa, R.: Estimation and choice of neighbors in spatial-interaction models of images. IEEE Trans. Information Theory 29:1 (1983) 60-72

[1983] Lebedev, D. S., Bezruk, A. A., Novikov V. M.: Markov Probabilistic Model of Image and Picture. Preprint: Inst. of Information Transmission Problems, Acad. Sci. USSR. VINITI (1983) [*In Russian*]

[1995] Li, S. Z.: Markov Random Field Modeling in Computer Vision. Springer (1995)

[1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E.: Equations of state calculations by fast computing machines. J. of Chemical Physics 21 (1953) 1087-1091

[1971] The Oxford English Dictionary (the Compact Edition). Vol. II (P-Z). Oxford Univ. Press (1971) 1165, 3274.

[1995] Pickard, R., Graszyk, C., Mann, S., Wachman, J., Pickard, L., Campbell, L.: VisTex Database. Media Lab., MIT, Cambridge, Mass. (1995)

[1993] Tuceryan, M., Jain, A. K.: Texture analysis. Handbook on Pattern Recognition and Computer Vision (C. H. Chen, L. F. Pau, P. S. P. Weng, Eds). World Publishing (1993) 235-276

[1959] Webster's New Int. Dictionary. 2nd Ed. Unabriged. Spingfield: Merriam Co. Publ. (1959) 2501, 2614.

[1986] Webster's Third New Int. Dictionary of English Language. Unabriged. Spingfield: Merriam-Webster Publ. (1986) 2267, 2366.

[1995] Winkler, G. Image Analysis, Random Fields and Dynamic Monte Carlo Methods. Springer (1995)

[1988] Younes, L.: Estimation and annealing for Gibbsian fields. Annales de l'Institut Henri Poincare 24:2 (1988) 269-294

[1994] Zalesny, A. V.: Homogeneity & texture. General approach. Proc. 12th IAPR Int. Conf. on Pattern Recognition. October 9-13, 1994, Jerusalem, Israel. Vol. 1. Los Alamitos: IEEE Comp. Soc. Press (1994) 592-594.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399